Electrophysiological indices of preparation and behavioural performance measures for episodic memory retrieval

Angharad Nia Williams

2016

A thesis submitted to Cardiff University for the degree of Doctor of Philosophy in Psychology

General abstract

The experiments in this thesis were designed to understand how and when preparation for episodic retrieval occurs, and what functions effective preparation enables. It has been proposed that in order to engage in episodic memory retrieval one must be in a certain cognitive state, known as retrieval mode, which ensures that stimulus events are treated as episodic memory cues. The experiments described below identify boundary conditions for when correlates of retrieval mode can be observed, and contribute new information about how retrieval mode contributes to episodic retrieval. This was accomplished via analysis of behavioural and Event-Related Potential (ERP) data.

The starting point for the work in this thesis was the observation that to date ERP correlates of retrieval mode have been obtained by comparing electrophysiological data when participants prepare to complete episodic or semantic memory tasks. Across a number of studies preparatory activity associated with the episodic task has been more positive-going at right-frontal scalp sites than preparatory activity associated with the semantic tasks. In these studies, however, the content of the memory being retrieved also varied across the tasks. In order to assess the possibility that the differences observed previously were due to differences between the task contents, Experiments 1 and 2 were designed to investigate the effect of equating the content of what was to be retrieved. In these experiments, during the study phase objects were presented either inside or outside an abstract outline of a building. At test, the objects from the preceding study phase were intermixed with unstudied objects, and participants were pre-cued on a trial-by-trial basis as to which task was to be completed. The episodic task required retrieval of the object location at study (inside/outside/new). In Experiment 1, the content of the semantic task was matched with the content of the episodic task. In Experiment 2, the content of the semantic task differed from the episodic task. In both cases there was no evidence for the pattern of activity associated with retrieval mode in previous studies. These results suggest that the inferences made previously about consistent neural signatures associated with preparation for episodic retrieval do not hold under certain circumstances.

Experiment 3 was designed to investigate whether the index identified previously is material-specific. This experiment was identical to Experiment 2, but words were used instead of pictures (as was the case in the previous published studies). There was also no evidence in

this experiment for the putative retrieval mode ERP index. In Experiment 4, the trial timings and the predictability of the task cue sequence were adjusted, with a view to establishing which design elements contributed to the pattern of null results observed in Experiments 1-3. This experiment was identical to Experiment 1, the exceptions being that the response-cue interval was shortened to 500ms (1200ms in Experiment 1) and the sequence of the preparatory cue indicating which task to complete was unpredictable. In this experiment there was a greater relative positivity following the episodic task cue on switch trials over right-frontal scalp sites. There was, however, no evidence for the generic retrieval mode ERP index on stay trials (the second trial of the same task), which is where the modulation was identified previously. These findings highlight the need to consider design factors to constrain explanations for when preparation for episodic retrieval is observed, and, consequently, to understand the benefits that it affords.

Additionally, in Experiments 1, 2, and 4, accuracy switch costs were evident. These have not been typically observed in previous studies, although reaction time costs have. In Experiments 5 and 6, the boundary conditions for these behavioural switch costs were investigated, with the findings indicating that the accuracy costs are restricted to switch trials and that for reaction time there are also costs associated with the general requirements to switch between tasks. These outcomes provide a starting point for understanding the boundary conditions for different kinds of switch costs in memory tasks.

Acknowledgements

Table of contents

1

3

List of tables

1. CHAPTER 1. General introduction to the control of episodic memory

1.1 Episodic memory retrieval

The focus in this thesis is on control over retrieval from episodic memory. This form of memory is defined as conscious memory for events in subjective space and time (Tulving, 1993), which distinguishes it from semantic memory which is defined as knowledge of objects and their interrelationships within the world (Tulving, 1993). One important determinant of episodic memory retrieval is what happens at the time of memory encoding (Craik & Lockhart, 1972; Morris, Bransford, & Franks, 1977; Tulving & Thomson, 1973). The focus here, however, is on processes that occur around the time of retrieval.

It has been proposed that the retrieval of episodic memories involves the interaction of a retrieval cue (generated internally, or a stimulus in the environment) and a memory trace (information represented in memory) (Rugg & Wilding, 2000; Schacter, Eich, & Tulving, 1978; Semon, 1921). This is known as the cue-trace interaction, and has been labelled 'ecphory' (Schacter et al., 1978; Semon, 1921). Many stimuli that we encounter on a daily basis could act as cues for certain memory traces. However, it has been argued that we do not prioritise these cues unless we are required to selectively retrieve relevant memories (Tulving, 1983). Thus, by one account, the additional explicit requirement that one should intentionally attempt to recall a particular event instantly influences whether episodic memory retrieval occurs (Tulving, 1983). The processing stages that guide the interaction between memory cues and internal memory representations have been described elsewhere as controlled memory retrieval operations (Mecklinger, 2010).

1.2 Cognitive control for episodic memory retrieval

Broadly defined, control processes are operations initiated intentionally in order to solve a particular task (Buckner, 2003). According to Schneider and Shiffrin (1977) cognitive control involves sequential processing steps, and these processes occur on occasions where automatic stimulus-response mappings cannot meet the needs of the task goal, and are subject to capacity limitations. Within the domain of memory, there is far more information available than can be accessed at any single point in time (Tulving & Pearlstone, 1966). Thus, central to intentional memory retrieval is controlled processing (Velanova et al., 2003).

Expanding on this account, there is the need to search, constrain, select, and monitor available information (Buckner, 2003; Velanova et al., 2003) in the service of task goals. For instance, relevant features related to an event may be retrieved in response to a retrieval cue, based on the retrieval objectives and intentions, and 'binding mechanisms' may integrate together these fragments of material (Mecklinger, 2010).

In one influential early account, Burgess and Shallice (1996) proposed a model for the relationship between control processing and the recall of details from one's own personal past. This model was based on evidence obtained from conversations about events from the personal past of healthy human volunteers (autobiographical recollection protocols). Different layers of control were noted in the analysis of the recollection protocols (Burgess & Shallice, 1996). Burgess and Shallice (1996) reasoned that the errors or lapses evident when healthy humans provide a running commentary of their recall process is related to the retrieval failures observed in patients with neurological problems. This analogy, and the breakdown of different components of their model, was used to explain patterns of performance in neurological patients who confabulate (Burgess & Shallice, 1996).

Confabulation is the production of narrative descriptions of events that have never happened (Burgess & Shallice, 1996; Schacter, Norman, & Koutstaal, 1998). Characteristically, confabulators are not aware that their narratives are false, and they are not intentionally produced (Burgess & Shallice, 1996; Schacter et al., 1998). Confabulation occurs in Korsakoff's Syndrome, or when individuals have lesions to the frontal lobes (Schacter et al., 1998). Confabulators typically draw on fragments of information from their personal past, but these episodes are commonly confused in time and space. Confabulation has been considered a deficit at the time of retrieval, because the content often concerns the time before the patient displayed problems (Burgess & Shallice, 1996; Moscovitch, 1989).

In an attempt to explain the phenomenon, Burgess and Shallice (1996) proposed an overarching three-stage process. This model was based on an earlier four-stage model by Norman and Bobrow (1979) which included retrieval specification (where a description of the target memory is specified, and verification criteria are produced), matching of target information with memory records, evaluation with reference to the verification criteria, and subsequent modification. Norman and Bobrow (1979) regarded retrieval of the memory record as an all-or-none protocol. Conversely, Burgess and Shallice (1996) suggested a more

flexible model, whereby the memory record may be broken down into smaller constituent parts which in turn may be confused with parts relating to other events.

Burgess and Shallice (1996) used the terms 'Descriptor, Editor, and Mediator' to describe their three operations. They reasoned that there must be processes that convert the demands of the retrieval goal into a form that can address the long-term storage systems. It was proposed that the Descriptor specifies the type of trace that would satisfy requirements: it defines the precise set of paths which correspond to aspects of specific events. From the autobiographical recollection protocols they obtained, and from the confabulations of neurological patients, there was evidence of failures in this stage. For example, both groups displayed conflation of memories (although more so in the neurological patients), and the incorporation of more recent input into old memories. Burgess and Shallice (1996) suggested that if there is too much of the store available then one is not able to effectively inhibit inappropriate parts.

The Editor was proposed as the detector of incompatibilities, evident in 'verbal checking' and pausing as the volunteers recalled information. These behaviours were noted to be lacking in the confabulators, in conjunction with their increased speed of response (Burgess & Shallice, 1996). The Mediator was proposed to resolve the incompatibilities detected by the Editor, via the use of problem-solving routines under strategic control (Burgess & Shallice, 1996). In healthy humans Burgess and Shallice (1996) noted evidence of reasoning processes, whereas for the confabulators reasoning errors were more abundant (especially for a certain kind of confabulation, where the narrative is completely bizarre to any outsider).

Moreover, from the recollection protocols with the healthy volunteers it was noted that retrieval does not often follow the order of the experienced event (Burgess & Shallice, 1996; Moscovitch, 1989). According to Burgess and Shallice (1996), the order of retrieval is dependent on the relative strength of connections. In addition, Burgess and Shallice (1996) noted that the assumption of 'schematisation' is relevant to particular aspects of confabulation, whereby there is evidence of the retrieval of a subset of elements corresponding to a particular event which is then generalised across events (known as 'summarisation').

13

Complementing this line of thought, Schacter et al. (1998) also provided a general integrative framework outlining the problems that the human memory system is required to solve in order to accurately retrieve details from the personal past. Schacter et al. (1998) observed that the retrieval of memories is not a literal representation of the past, and proposed that instead it involves a constructive process. According to Schacter et al. (1998), it is this constructive process that is on occasions prone to errors, distortions, and illusions of memory. In fact, it was noted that mere contemplation of an event can result in the construction of a detailed and vivid representation, which can be difficult to tell apart from a stored representation of an event that was actually experienced (Schacter et al., 1998).

Schacter et al. (1998) proposed the Constructive Memory Framework (CMF), which at the time of retrieval includes the initial formation of a retrieval description that is sufficiently focused to query memory, followed by later post-retrieval monitoring and verification stages. Like Burgess and Shallice (1996), Schacter et al. (1998) provided a review of ideas for how and why different memory distortions occur. Different component processes were discussed and phenomena of constructive memory from healthy humans and patient case studies were provided to evidence this. For instance, mis-combinations (or memory conjunction errors, Reinitz, Lammers, & Cochran, 1992) are evident in studies of healthy participants, likewise evidence from confabulators demonstrates that often specific information is provided but is from an inappropriate context (Schacter et al., 1998).

Specifically, at retrieval, a large part of the CMF is concerned with the need for 'focusing'. Schacter et al. (1998) proposed that retrieval cues need to be specific to allow correspondence with one single event. When this is not the case, multiple potential possibilities will be accessed and compete which will result in poor memory for 'source specifying' details (M. K. Johnson, Hashtroudi, & Lindsay, 1993) which differentiate them (Schacter et al., 1998). This idea corresponds to the Descriptor stage identified by Burgess and Shallice (1996). Other processes at retrieval that, according to Schacter et al. (1998), demonstrate the constructive nature of human memory include pattern completion and criterion setting (which resonate with the Editor and Mediator stages put forward by Burgess & Shallice, 1996).

Consistent across these models of episodic memory is the description of control processes carried out prior to, during, and as a consequence of retrieval (Wilding & Herron, 2006). It has been widely assumed that these processes determine the mnemonic information that is made available, assessed, and then used to guide behaviour (Burgess & Shallice, 1996; Rugg & Wilding, 2000; Schacter et al., 1998; Wilding & Herron, 2006). In general terms it has been suggested that processes engaged before the cue-trace interaction could guide the nature of that interaction (and this can come about in various ways), and the processes engaged during or after the interaction could operate on the consequences of the interaction depending on task requirements (Wilding & Herron, 2006). There is evidence from neuropsychological patients of breakdown in successful memory processing before the cue-trace interaction (Wheeler, Stuss, & Tulving, 1997), and evidence of problems at later post-retrieval stages (Ranganath & Knight, 2002).

This general observation has been captured in a more recent broad framework by Rugg and Wilding (2000), who provided a four-fold classification of sub-processes involved in controlled memory retrieval. These are: retrieval mode, retrieval orientation, retrieval effort, and retrieval success, and each will be briefly summarised in turn.

1.2.1   Retrieval mode

Tulving (1983) proposed a prerequisite for successful episodic memory retrieval to occur. Based on casual observation, he suggested that one must be in an appropriate cognitive state or task-set for the cue-trace interaction to ensue. As an example of this idea, he discussed the instance of an unexpected encounter with a friend. Tulving (1983) stated that our initial thoughts would be why the friend is in that place at that particular time, rather than the retrieval of memories about previous events involving that friend. Tulving (1983) further offered the concept of retrieval mode to describe the state that one must be in to engage in episodic memory retrieval. It has been suggested that this cognitive set allows inputs from the environment to be treated as retrieval cues for past events, permitting the 'mental time travel' enabling one to think in subjective time and space (Nyberg et al., 1995; Tulving, 1983; Wheeler et al., 1997).

This notion of the initiation of a task-set for episodic retrieval resembles the initial stages of the theoretical models mentioned previously. The idea of early specification processes (Schacter et al., 1998), and the idea of a Descriptor based on protocol evidence from healthy humans (Burgess & Shallice, 1996), corresponds with Tulving's general observations and the retrieval mode concept (Tulving, 1983).

1.2.2   Retrieval orientation

Rugg and Wilding (2000) developed the concept of a more constrained task-set than retrieval mode, which they described as a retrieval orientation. This is the more specific task-dependent form of processing applied to a retrieval cue (Rugg & Wilding, 2000). The purpose of retrieval orientations is to constrain the processing of retrieval cues based on the specific episodic task demands (Rugg & Wilding, 2000).

1.2.3   Retrieval effort

The concept known as retrieval effort refers to the levels of processing resource utilised during attempts to retrieve (Rugg & Wilding, 2000). This is proposed to differ for tasks varying in difficulty, and may reflect changes in the levels of engagement of the same set of processes, and/or the engagement of qualitatively different resources.

1.2.4   Retrieval success

Operations that have been paid the most attention to date in psychology and cognitive neuroscience are those that fall under the category of retrieval success. Any process that depends on ecphory comes under this description (Rugg & Wilding, 2000). Thus, these are the processes that are engaged during and as a consequence of the cue-trace interaction (Wilding & Herron, 2006). For instance, the acontextual sense that something has been previously experienced is known as familiarity-based recognition (Jacoby, 1991; Mandler, 1980; Woodruff, Hayama, & Rugg, 2006; Yonelinas, 2002). Additionally, according to dual-process frameworks there is another process, commonly known as recollection, which supports the retrieval of contextually specific information such as where or when events occurred (Jacoby, 1991; Mandler, 1980; Woodruff et al., 2006; Yonelinas, 2002). It has further been suggested that there are processes occurring later (Wilding & Rugg, 1996), perhaps for post-retrieval monitoring and evaluative purposes (but see Hayama, Johnson, & Rugg, 2008), some of which also fall under the broad category of retrieval success.

1.2.5 Summary

In summary, difficulties with retrieving memories may emerge due to failures at one (or more than one) of many goal-directed processing stages, and evidence from individuals with selective brain damage is consistent with the proposal of a multi-stage component model. However, the evidence base for processing of this kind has also been developed from patterns of neural activity recorded from healthy humans, and a review of previous studies relevant to this issue follows in section 1.4. This is the main approach taken in the experiments reported in this thesis: investigations of neural correlates recorded from normal functioning individuals. In advance of this description of imaging data obtained in memory studies, the immediately following section contains an overview of the cognitive and brain imaging task-switching literature, along with consideration of the key assumptions that are made in cognitive psychology and cognitive neuroscience studies regarding isolation and definition of processes of interest. The emphasis on task-switching arises because each of the experiments in this thesis requires participants to alternate between completion of different tasks. The broad rationale for this approach is that it provides a means of manipulating conditions under which variations in the engagement of cognitive sets like retrieval mode can be observed, as will be outlined in greater detail later in this thesis.

1.3 Cognitive sets and task-switching designs

During task-switching experiments participants are required to switch frequently between two or more tasks. Other labels that have been used for experimental designs where participants are required to switch between tasks are set-shifting and cognitive flexibility. The purpose of task-switching manipulations is to provide a means of assessing the properties and influence of task-sets and cognitive control operations. In psychology experiments, participants are frequently required to read task instructions and respond accordingly during the experiment. Thus, participants chain together and configure appropriate response options via processes linking sensory analyses to motor output (Rogers & Monsell, 1995). According to Rogers and Monsell (1995), it is this categorisation and assembly that constitutes the cognitive task-set. That is, a task-set is a cognitive state whereby one has formed an effective intention to perform a specific task (Rogers & Monsell, 1995). In terms of implementation, a task-set can be considered to have been adopted when mental resources are organised in order to accomplish a particular task, depending on the appropriate input (Monsell, Sumner, & Waters, 2003), and these have been described as 'internal control settings' (Mayr & Kliegl,

2000). In this same vein, Monsell (2003) refers to the act of introspection, where one can put oneself into a specific prepared state whereby subsequent sequences of operations unfold as a 'prepared reflex'. This last notion is important because it entails an assumption that at least some processes will be engaged relatively automatically once a set has been adopted. A similar emphasis can be seen in consideration of how control over behaviour can be split into separable components, with one being between endogenous and exogenous control (Monsell et al., 2003; Rogers & Monsell, 1995). Endogenous control is the internally driven intentional component of task control. Exogenous control is the habitual processing afforded certain groups or types of stimuli, with Stroop phenomena being one example of this kind (Stroop, 1935). A third component, described as autogenous priming, has also been suggested (Monsell et al., 2003). This priming reflects the recency and frequency of task-set use (see also Norman & Shallice, 1986).

If one is to assess the properties of task-sets, it is necessary to have a set of paradigms to investigate them. Critical criteria the paradigms must meet are the ability to assess the influences of sets under conditions where they might be engaged to greater or lesser degrees, and/or when there is the possibility of interference between different sets. Task-switch paradigms meet these criteria, and an influential early approach was developed by Jersild (1927). In this paradigm the participant either (i) switches task from trial-to-trial, or, (ii) completes the same task on multiple trials. This basic paradigm has been developed into several variations in the last 20 years. Examples of these variations in task-switching designs include the alternating-runs paradigm, the task-cueing paradigm, and the intermittent-instruction approach (Monsell et al., 2003).

In each of these designs where frequent switches are required, the first trial of a task is commonly referred to as a *switch* trial, and a second successive trial of the same task as a *stay* trial (sometimes referred to as nonswitch, switch+1 or repeat trials: Monsell, 2003; Morcom & Rugg, 2002; Werkle-Bergner, Mecklinger, Kray, Meyer, & Düzel, 2005). Performance measures are commonly compared for switch and stay trials. In the general task-switching literature, reaction time differences between switch and stay trials are referred to as the reaction time switch cost (Wylie & Allport, 2000). Error rates and other accuracy measures are also frequently assessed, and in general performance is poorer on switch than on stay trials.

Within the task-switching literature, the primary research focus has been on factoring out the relative contributions of the array of processes that could contribute to the behavioural costs that are typically observed. The costs have been proposed to reflect numerous executive control processes such as: shifting attention, the retrieval of goal relevant information, reconfiguration of the task-set, and inhibition of interference from the preceding task-set (Monsell, 2003). In addition, task-priming processes have attracted additional interest (Logan & Bundesen, 2003).

It is clear that costs arise for multiple reasons, and are due to processes operating at discrete processing stages (for reviews see Kiesel et al., 2010; and Vandierendonck, Liefooghe, & Verbruggen, 2010). Brain imaging approaches have been employed to investigate several of these processes, since they provide opportunities for assessing properties of sets that are not available via behavioural measures alone (for a review see Jost, De Baene, Koch, & Brass, 2013). One issue to which ERP and functional Magnetic Resonance Imaging (fMRI) data have contributed is the question of the role of task-set inertia (interference from the previous task) in switch costs (Allport, Styles, & Hsieh, 1994; Düzel et al., 1999; Wylie & Allport, 2000). The ERP data from Evans, Herron, and Wilding (2015) demonstrated more direct evidence for task-set inertia: the ERP effects related to the previous task were larger on switch than on stay trials. Moreover, Evans, Herron, et al. (2015) also demonstrated that the magnitude of the ERP effects was related to the behavioural switch cost: as the task-irrelevant activity reduced so did the reaction times. In addition, exploiting the different activation profiles during the processing of faces and words, Yeung, Nystrom, Aronson, and Cohen (2006) provided fMRI data demonstrating that there was increased activation of task-irrelevant information immediately after a task-switch. They also demonstrated that the magnitude of the fMRI activation was correlated to the reaction time switch cost.

A second substantive contribution is in relation to discussions over cue- versus task-switching. Logan and Bundesen (2003) adapted the task-cueing procedure, so that each task could be indicated by two different cues. They observed that even if the task remains constant, cue-switches cause substantial costs (for discussion see Altmann, 2007; Jost et al., 2013; and Jost, Mayr, & Rösler, 2008). However, critically, there is electrophysiological and neuroimaging evidence consistent with a functional dissociation between these two kinds of switches (Jost et al., 2013). Jost et al. (2008) demonstrated distinct features of ERP data

19

related to task- and cue-switches, suggesting distinct processing stages and separate neural generators. On this basis, Jost et al. (2008) concluded that task-switches involve a unique set of mental processes that are distinct from cue-switches. These outcomes substantiated earlier work. For example, Nicholson, Karayanidis, Bumak, Poboka, and Michie (2006) also demonstrated ERP differences when switching between task-sets and task-cues. In addition, in an fMRI study by Brass and von Cramon (2004) activation in brain regions linked to preparation was associated with task-switching rather than cue-switching.

These examples from the brain imaging literature provide support for Monsell's observation that task-switching designs lend themselves well to the subtractive methodology of neuroimaging and electrophysiology (Monsell, 2003). Although Monsell went on to qualify this claim by noting the complexities of the processing operations that will contribute to switch costs, his observations are a useful departure point for considering fundamental assumptions that underpin cognitive psychology and cognitive neuroscience.

The subtractive methodology is based historically in the field of 'mental chronometry' (Posner, 1978) and the initial work of Donders (1868/1969) was pivotal. Donders designed a method to provide an estimate of the duration necessary for a particular mental operation. This method is known as the subtraction method. In Donders' original experiment the left or right foot of human subjects were electronically stimulated. The participant was required to respond with the hand that was on the same side of the body as the foot that was stimulated. There were two conditions: simple and choice reaction time. For the simple reaction time condition, the stimulated foot was constant throughout a block of trials. For the choice reaction time condition, the stimulated foot varied trial-to-trial. By subtracting the simple reaction time from the choice reaction time, Donders' reasoned that one was able to reveal the time required to decide which side had been stimulated, thus, for establishing response selection ('the action of the will') of the left or right side. Simple reaction time and choice reaction time are described as 'stimulus discrimination' and 'response selection' in more current terminology (Coles, Smid, Scheffers, & Otten, 1995).

In general terms this approach to cognitive psychology allows one to make inferences about mental processes from measures of reaction time. It enables one to answer detailed questions about the structure and function of a covert system from measures of overt behaviour (Coles et al., 1995), which is the basic-level challenge for cognitive psychology.

Including but extending beyond reaction time measures, a common approach is to design tasks such that cognitive processes can be isolated or separated via experiment conditions that differ only in the critical process(es) of interest. This approach is central to the cognitive paradigm in which it is assumed that cognition can be detailed as a set of related (serial and parallel) abstract processes that act over representations (for challenges to this view from a connectionist perspective see Fodor & Pylyshyn, 1988; McClelland, 1989; and Smolensky, 1987).

The idea that the only difference between one condition and another is in the 'pure insertion' of a processing stage has been challenged. For reaction times, the assumption that the incremental effect on reaction time is strictly additive did not hold in all circumstances (Coles et al., 1995). Motivated by consideration of the adoption of the subtraction methodology in functional imaging experiments (where the assumption is that processes can be isolated by subtracting neural activities between conditions that differ only in the process of interest), Friston and colleagues have emphasised the fragility of the assumption that specific processes will operate in the same way irrespective of the context in which they operate (Friston, 2005). If this assumption does not hold, then there will be challenges with isolating the same process via different experimental manipulations, but there are solutions, such as a parametric approach, as well as the observation of consistency of outcomes (which might be neural and/or behavioural) across manipulations intended to isolate the same phenomena.

The assumption that, via appropriate experiment designs, it is possible to isolate processes of interest underpins the work in this thesis. The approach here involves exploiting the temporal resolution of ERPs to provide insights into processing stages that from a purely behavioural perspective can only be inferred indirectly. More specifically, the approach involves recording neural activity while participants prepare to make different kinds of judgments to assess cognitive sets that might be engaged during preparation for retrieval. Understanding the functional significance of these sets depends critically, of course, on linking neural measures to measures of behaviour, and this link is maintained throughout this thesis, as it has been in the existing literature on preparation for retrieval that is described in the next section.

1.4 Neural correlates of retrieval sets

Brain activity recorded from healthy individuals has been used to investigate retrieval mode in a number of studies. To identify this cognitive state with a neural measure, Rugg and Wilding (2000) proposed that there are three criteria that must be met. Firstly, the neural activity should be time-locked to the onset, and maintained for the duration, of an episodic retrieval task. Secondly, a neural correlate should be revealed when contrasting the activity observed during episodic retrieval tasks to that observed during non-episodic tasks. Finally, the activity should not vary across different episodic retrieval tasks.

1.4.1 Haemodynamic imaging studies

1.4.1.1 Positron Emission Tomography (PET)

One of the first studies in which retrieval mode was investigated using brain imaging was reported by S. Kapur et al. (1995). PET was used to measure regional Cerebral Blood Flow (rCBF) while three tasks were completed in separate blocks. There was a semantic control task, and two recognition memory tasks which consisted of high and low target recognition conditions in which the number of old stimuli (targets) varied. Neural activity in right Pre-Frontal Cortex (PFC) was equivalent for both recognition conditions but higher than for the semantic condition. Neural activity in the right medial parietal region of the precuneus was greater for the high than the low recognition condition. S. Kapur et al. (1995) suggested that the prefrontal activation indexed retrieval mode because it differentiated between the episodic and semantic tasks, but was insensitive to the amount of content retrieved (which was assumed to be greater in the high than the low retrieval condition because of the differences between the numbers of old and new stimuli). The precuneus activation was linked to successful retrieval because it was sensitive to these numerical differences.

These interpretations are broadly consistent with those offered in earlier studies highlighting the involvement of the PFC (right more so than the left hemisphere), during episodic memory retrieval (Grasby et al., 1993; Shallice et al., 1994; Squire et al., 1992; Tulving, Kapur, Craik, Moscovitch, & Houle, 1994; Tulving, Kapur, Markowitsch, et al., 1994). Moreover, these interpretations are consistent with models based on neuropsychological evidence and theory, in which memory control processes are considered and have been linked with the integrity of the frontal cortices (Moscovitch, 1992).

Another relevant experiment was reported by Nyberg et al. (1995). At study participants heard concrete nouns spoken one at a time. In one phase participants had to classify the voice as male or female, and in another participants had to classify the object denoted by the noun as living or non-living. Regardless of how the information was encoded, during three visual old/new recognition phases (one containing only new items) there was increased activation in the right PFC (as well as the left anterior cingulate and cerebellum), which was absent in a phase when participants were required to silently read the words presented to them. Nyberg et al. (1995) proposed that this activation pattern reflected retrieval mode because it was observed independently of retrieval success, which could not have occurred for blocks containing only new words. Moreover, here, the right prefrontal activation was observed for information encoded aurally, suggesting that the activation identified previously and described above (S. Kapur et al., 1995) was not a modality-specific effect. In addition, analogous to the design by S. Kapur et al. (1995), it was expected that retrieval success would vary in a graded manner, with higher levels in the condition with meaning (living/non-living), than the voice condition (male/female), and absent during the new item only phase. Contrasts between these conditions revealed regional brain activation associated with varying levels of retrieval success, depending on how the information was encoded. Each encoding condition was associated with a distinctive pattern of rCBF: the meaning condition with increased activation in left temporal cortex, and the voice condition involved regions in right orbital frontal and parahippocampal cortex (Nyberg et al., 1995).

A subsequent PET study focused on the brain regions supporting recall and recognition (Cabeza et al., 1997). Based on frontal lesion data demonstrating reliably impaired recall but not recognition performance, the researchers predicted that activation in the PFC would dissociate according to whether participants were recalling or recognising previously studied items (see Janowsky, Shimamura, Kritchevsky, & Squire, 1989; Jetter, Poser, Freeman Jr, & Markowitsch, 1986; but also Wheeler, Stuss, & Tulving, 1995). However, in contrast to a baseline reading task both cued recall and recognition memory were associated with increased PET activation in the anterior cingulate and right PFC. The task invariance of this activation was interpreted as further evidence consistent with the view that right PFC supports retrieval mode (Cabeza et al., 1997).

Additionally, Lepage, Ghaffar, Nyberg, and Tulving (2000) reported a small-scale meta-analysis of data from four PET studies (Düzel et al., 1999; S. Kapur et al., 1995;

23

Nyberg et al., 2000; Nyberg et al., 1995). Across the experiments, regions that were differentially activated for tasks involving episodic memory retrieval relative to non-episodic tasks, and independent of successful retrieval, were proposed to support or contribute to retrieval mode. Lepage et al. (2000) reported that three right prefrontal sites, two left prefrontal sites and a region in the anterior cingulate gyrus had this activation profile.

Lepage et al. (2000) also presented another review of results obtained from a survey of other episodic retrieval activations reported in the literature (Cabeza & Nyberg, 2000) and a database of studies used in a previous meta-analysis (Lepage, Habib, & Tulving, 1998). Forty PET/fMRI studies were selected, all of which included an episodic and a non-episodic retrieval task. Lepage et al. (2000) classified Talairach activation peaks (Talairach & Tournoux, 1988) from these studies as 'retrieval mode matches' if their vector distance was within 10mm from a retrieval mode site identified in the initial four-study meta-analysis. In addition, peaks within 10-16mm of any of the six retrieval mode sites were classified as 'near mode matches'. Of these 40 studies, 32 were classed as having at least one 'retrieval mode match', and an additional five were classed as having at least one 'near mode match'. Of the remaining three studies, two demonstrated right-frontal activation but this was not within 16mm of any of the predetermined areas (Bäckman et al., 1997; Schacter et al., 1996), and the other was an investigation of face processing (Kapur, Friston, Young, Frith, & Frackowiak, 1995).

The studies reported thus far fit some of the identification criteria for a signature of retrieval mode as proposed by Rugg and Wilding (2000). In the studies listed above, engagement of the right PFC was generally evident during episodic retrieval, regardless of the episodic task, and was not dependent on the nature of the retrieval cue. However, the claim that these studies demonstrate a generic functional anatomic correlate of retrieval mode can be challenged. Across individual haemodynamic imaging studies, reports of the involvement of specific fronto-polar/frontal regions have varied. Attempts to identify consistent activation areas have been made using meta-analytical techniques. However, these analyses have also demonstrated that there is some variability in the frontal regions identified across tasks (Lepage et al., 2000).

Moreover, Rugg, Fletcher, Frith, Frackowiak, and Dolan (1996) and Rugg et al. (1998) manipulated the density of old items in test lists and observed changes in right-frontal

activation with density. Rugg et al. (1998) concluded that activation of the right anterior PFC reflects item-related memory 'monitoring' operations during episodic retrieval. Alongside these inconsistent results, a fundamental observation is that item-related activity cannot be extracted with PET because of the need to average activity over many successive trials to obtain a measurable signal. This means that using PET it is not possible to infer whether a given time-averaged measure reflects a signal that was maintained over the time period of interest, or whether it varied - for example, with the measure reflecting the summation of activity elicited by the items falling within that time period. For this reason, the outcomes of PET studies, in isolation, can provide only limited evidence for a neural signature considered to reflect retrieval mode. Düzel et al. (1999) noted this caveat, and complemented a PET study with Direct-Coupled (DC) ElectroEncephaloGraphy (EEG) and ERP methods (see also Düzel et al., 2001). Acquiring these kinds of data enables concurrent investigation of transient neural activity and sustained low-frequency neural activity over several seconds (Rockstroh, Elbert, Canavan, Lutzenberger, & Birbaumer, 1989).

Düzel et al. (1999) monitored neural activity whilst participants completed an episodic retrieval task (old/new recognition memory) and a 'baseline' semantic task (living/non-living categorisation of items). The task-related activations differed in the form of a sustained low-frequency DC shift which was more positive-going for the episodic task in comparison to the semantic task, and most prominent over right-frontal scalp. In this experiment words were presented one at a time in mini 4-item blocks, and changes in activity with task were evident from the onset of the initial task cue, increased until the second word in the block was presented and were sustained for the remaining length of the block. These findings were interpreted as evidence for the development and/or maintenance of retrieval mode (Düzel et al., 1999).

The same experiment was run while PET data were acquired, and greater activation for the episodic than the semantic task was observed in the right-frontal and posterior cingulate cortex. The right-frontal PET activation converged with the outcome of source localisation of the ERP data. Furthermore, source analysis of the ERP data mapped onto the right PFC areas identified in previous PET studies. The findings suggest that the time aggregated PET measures from right PFC were a reflection of a continuous signal, rather than the summation of several activations linked to the items presented during a task.

1.4.1.2 Functional Magnetic Resonance Imaging (fMRI)

Like PET, fMRI is a haemodynamic imaging technique, but does not require intravenous administration of a radio-active tracer. Rather than measuring rCBF, fMRI detects changes in regional blood oxygenation (Toga & Mazziotta, 2002). As with PET, blocked fMRI procedures confound item- and task-related operations because of the time integrated measure of activity. With event-related fMRI, however, item-related activity can be identified, and in addition there are means of separating item- and task-related activity from the same data set (Donaldson, Petersen, Ollinger, & Buckner, 2001; Wilding, 2001).

The neural activity related to specific 'item' and 'state' related processes was separated for the first time using fMRI by Donaldson et al. (2001), with a mixed 'blocked and event-related' procedure (see also Chawla, Rees, & Friston, 1999; and Otten, Henson, & Rugg, 2002). Donaldson et al. (2001) characterised 'task' related processes as 'state' related processes. State related processes (such as retrieval mode) were regarded as the operations associated with ongoing goals. Whereas processes associated with individual test items remained characterised as 'item' related processes, in this case related to the recovery of specific information from memory (Donaldson et al., 2001).

Donaldson et al. (2001) asked participants to complete alternating blocks of continuous fixation and a recognition memory task. Importantly, within the recognition memory task blocks, variable durations of fixation trials (jitter) were introduced between sequential memory trials (Donaldson et al., 2001). This approach allowed for item-related activation patterns to be identifiable across time (event-related), and statistically removed from what was determined as ongoing state-related activity.

Activation in distinct brain regions was associated with either state or item processing periods, providing support for the theoretical claim that a combination of state and item related processes contribute to recognition memory (Donaldson et al., 2001). Using three separate contrasts, three separable effects were revealed: (i) a retrieval success effect was evident in response to correctly recognised old items (in comparison to correctly rejected new items) and located in medial parietal, lateral parietal, and anterior left frontal cortex; (ii) a transient effect was evident in response to recognition test items regardless of item-type, revealing a network of areas commonly activated by visually cued recognition memory

(visual, frontal, and parietal regions); (iii) a sustained effect was also evident when activation during the memory task periods was contrasted with activation during the fixation block periods (Donaldson et al., 2001). However, it is noteworthy that here the sustained task index was not localised to the right PFC. A network of regions was revealed, and, although all were located in the frontal cortex, five out of seven were left lateralised, one was within a medial region, and only one was located in the right hemisphere (the right frontal operculum).

Following Donaldson et al. (2001), Velanova et al. (2003) investigated specific functional-anatomic correlates of sustained and transient processes during memory retrieval. Importantly, the findings in Velanova et al. (2003) corroborated claims that regions within frontal cortex dissociate on the basis of their temporal profiles. Here, right fronto-polar cortical regions were associated with sustained control processes, proposed to potentially underlie retrieval mode or an attentional set (Velanova et al., 2003). More posterior prefrontal regions were associated with transient individual retrieval attempts. Velanova et al. (2003) concluded that the use of a mixed blocked and event-related design, and the demonstration that right fronto-polar contributions to retrieval include a sustained component, facilitated the reconciliation of discrepant findings that arose when contrasting blocked-trial paradigms with event-related ones (for a review of blocked studies see Buckner, 1996; and Nyberg, Cabeza, & Tulving, 1996; for a review of event-related findings see Rugg & Henson, 2002; and see event-related study by Schacter, Buckner, Koutstaal, Dale, & Rosen, 1997). Moreover, Velanova et al. (2003) stated that they did not anticipate that involvement of the fronto-polar cortex was specific to episodic retrieval, but more broadly to all tasks demanding certain kinds of high control (such as controlled recognition, but also other working memory and decision tasks) (but see Buckner, 2003, for a different view).

In related work Simons, Gilbert, Owen, Fletcher, and Burgess (2005) and Simons, Owen, Fletcher, and Burgess (2005) investigated potential contributing factors to account for differential regional recruitment across studies. Using event-related fMRI, Simons, Gilbert, et al. (2005) employed a retrieval cue manipulation as a means of inferring the stage of the retrieval process that regions may be involved in. On all trials participants were cued to prepare for episodic retrieval. On some trials, a test item (retrieval cue) was then displayed and a memory judgment was required. On other trials participants were only required to press a button corresponding to a number presented on the screen in place of a retrieval cue. The authors reasoned that activity associated with a sustained retrieval mode would be evident

27

regardless of whether retrieval searches (and other additional processes that occur subsequently in response to the retrieval cue) were carried out.

Activation of the lateral anterior PFC was evident in both conditions, suggesting that the anterior PFC is involved in processes occurring before the instigation of a retrieval search. By contrast, activation in the medial anterior PFC was evident only when a subsequent search was carried out (Simons, Gilbert, et al., 2005). A time-course analysis was also conducted, and indicated that activation in the lateral anterior PFC peaked earlier than activation in the medial anterior PFC. From this it was inferred that lateral regions of the anterior PFC have more involvement in the specification of retrieval strategies, which might include retrieval mode, while medial regions have more involvement in retrieval search, and monitoring and verification stages (post-retrieval) (Simons, Gilbert, et al., 2005).

1.4.1.3 Summary

In summary, a number of different data points suggest that two separable classes of processes can be identified using neural measures because of the presence of sustained and transient patterns of neural activity. Setting aside issues of localisation, the findings provide broad support for the concept of retrieval mode. However, none of the studies fully meet the criteria suggested by Rugg and Wilding (2000); either because the design, the neural measure, or both do not permit particularly strong claims.

1.4.2   Real-time recording studies: ElectroEncephaloGraphy (EEG)

1.4.2.1   Background assumptions

Real-time temporal resolution is critical when trying to delineate the complicated interplay of processes that contribute to episodic memory retrieval. Distinct computations have overlapping time courses; thus separating them temporally is especially important to allow characterisation of each role. As is evident from the preceding sections, separation is not so straightforward whilst using haemodynamic imaging measures, because they lack the necessary temporal precision to distinguish clearly between the processes engaged before a retrieval cue is presented, and those that are engaged subsequently in light of the retrieval cue (Wilding & Ranganath, 2011). This is because the haemodynamic methods measure neural activity indirectly, via downstream changes in regional blood flow or oxygenation which are sluggish in nature (Rugg & Wilding, 2000).

The temporal resolution of ERPs means that it is possible to separate and index the activity related to distinct operations in real-time. The processes engaged before a retrieval cue is presented, during the early specification stages, are also known as preparatory processes (Wilding & Herron, 2006). Designs that enable separation of the activity related to this timeframe from the activity related to subsequent processes occurring after a retrieval cue are of note here. The focus of the electrophysiological studies reported in this thesis is on the neural activity during this preparatory period, prior to the presentation of test items to which judgments are required.

1.4.2.2   Preparatory task cue ERPs

Morcom and Rugg (2002) utilised a task-switching design and ERP measures to distinguish between activity that was evident prior to the onset of a retrieval cue and the activity related to the cue itself. In this study, the same episodic and semantic tasks were used as in the Düzel et al. (1999) PET and ERP blocked task investigations (see earlier ***Positron Emission Tomography (PET)*** section for further details). However, here, the task-switching test design enabled separation of the activity on a trial-by-trial basis.

The test followed 5-10 minutes after an initial study phase which involved participants incorporating each presented word into a short sentence. Each trial in the test began with a pre-item preparatory task cue, indicating that participants were to prepare to

complete either an old/new recognition memory task, or an animacy judgment task. Following each task cue a test word was displayed and responses were subsequently required via key press before the next trial began. The focus was on the neural activity following the onset of the preparatory task cues, which signalled that participants were to subsequently make either the episodic or semantic memory judgment. This preparatory task cue varied pseudo-randomly trial-by-trial (restricted to a maximum of three of the same kind in a row).

In contrast to the semantic (animacy) task, a sustained relative positivity followed the preparatory cue for the episodic task. The effect was largest over sites near the midline at central and right-frontal recording locations. This relative positivity started 500ms after the onset of the cue and lasted until the test item was displayed 1600ms later. Moreover, this positivity did not emerge until, and was only reliable on, the stay trials (Morcom & Rugg, 2002). Based on these findings, Morcom and Rugg (2002) proposed that it takes at least one trial of a task before an episodic task-set can be adopted. This finding is consistent with others in the task-switching literature, which suggest that the complete adoption of a task-set, after switching from another task, takes at least one complete trial (Monsell, 2003). Düzel et al. (1999) referred to this phenomenon as 'neurocognitive inertia'.

The scalp distribution of the effect observed by Morcom and Rugg (2002) was broadly similar to that reported by Düzel et al. (1999), and they proposed that this effect was linked to retrieval mode - a reasonable conclusion given the earlier findings and the fact that the same tasks were used in both experiments. Because the same tasks were used, however, the possibility remained that the findings would not generalise to other episodic memory tasks, which would be a requirement if the effects identified did indeed index retrieval mode (Rugg & Wilding, 2000).

Herron and Wilding (2004) employed a task-switching paradigm using different retrieval tasks to those used by Düzel et al. (1999) and Morcom and Rugg (2002). Potential preparatory ERP indices of retrieval mode and retrieval orientation were concurrently investigated. Herron and Wilding (2004) compared the ERP activity whilst participants were preparing for two episodic retrieval tasks and a semantic memory task. For one of the episodic tasks participants had to retrieve the location of the item from a study phase (left or right side of the monitor screen). For the other episodic task, participants had to recall the encoding task (operation) which they completed for that item during a study phase (an

animacy or pleasantness judgment). For the semantic task participants were required to state whether the object denoted by the word could move of its own accord. In the test phase participants were pre-cued trial-by-trial as to which of these tasks was to be completed, and each cue-type was presented for two consecutive trials before requiring a switch to one of the other tasks. So, while the number of trials before a switch was predictable the task that would be switched to was not.

Herron and Wilding (2004) found that whilst preparing to retrieve information about either the location or operation task, ERP activity was significantly more positive-going (following the task cue from 800-1900ms) at right-frontal electrodes than when preparing for semantic retrieval. Consistent with the findings of Morcom and Rugg (2002), significant differences were evident only on stay trials. It was proposed that this right-frontal activity reflected the adoption and/or maintenance of retrieval mode (Herron & Wilding, 2004). In addition, the fact that it was observed for different episodic tasks meant that it met one of the criteria for an index of retrieval mode that was not demonstrated by Morcom and Rugg (2002) or Düzel et al. (1999): episodic task independence (Rugg & Wilding, 2000).

Moreover, differences were also evident between the ERPs evoked by the two episodic task cues, primarily at *left* hemisphere anterior and central locations and again on stay trials only (800-1900ms). More positive-going ERPs were evident when preparing to recollect conceptual information (content from the operations task) in comparison to perceptual information (the location task). Thus, the results provided additional evidence for task-specific retrieval sets (retrieval orientations) depending on the kind of episodic task that was to be completed.

In terms of behaviour, Herron and Wilding (2004) reported that reaction times were slower on switch trials than stay trials. However, accuracy was not influenced by the switching manipulation: which was also the case in the experiment conducted by Morcom and Rugg (2002). Based on these findings, Herron and Wilding (2004) suggested that the adoption of a retrieval set may be beneficial for maximising the efficiency of search operations, or the efficiency with which retrieved information is processed, rather than influencing what is in fact recovered from memory.

31

During a series of three experiments, Herron and Wilding (2006b) also investigated preparatory task-related ERPs. A task-switching design was utilised in two of the experiments to enable comparisons across switch and stay trials. The third experiment was a blocked experimental design. The two task-switching experiments differed in that in one the cue sequence was predictable (always two consecutive trials of each task) and in the other it was unpredictable (one, two, or three consecutive trials of each task). There were two episodic tasks: in one task participants had to recall the location of items from a prior study phase, and in the other participants had to recall the encoding task (operation) they completed for that item during the study phase. Preparatory task-related ERP indices were evident, and qualitatively similar, only in the two experiments where there were frequent switches between different retrieval tasks. From this it was inferred that this preparatory measure of brain activity reflects processes important for the initial adoption and configuration of a retrieval set (Herron & Wilding, 2006b), because it was absent in the experiment where presumably the appropriate task-set could be maintained from trial-to-trial. This effect was largest at left fronto-temporal sites, and the fact that this distribution is not equivalent to the one linked to retrieval mode is consistent with the claim that different task-sets are adopted according to the kind of episodic retrieval that is required (Herron & Wilding, 2004).

However, this preparatory index was evident on switch trials only. Previous indices of preparation for episodic memory retrieval had only been identified on stay trials (Herron & Wilding, 2004; Morcom & Rugg, 2002). It was proposed that switching between different episodic tasks, rather than in and out of completing an episodic task, could account for this discrepancy (Herron & Wilding, 2006b; J. D. Johnson & Rugg, 2006). Herron and Wilding (2006b) observed that participants in the switching experiments had to reconfigure to retrieve different contents within the episodic memory system (switch between two episodic tasks). Switching in the previous studies, however, may have placed greater demands upon set reconfiguration, as participants switched in and out of making episodic memory judgments (switch between an episodic and a semantic task). Thus, importantly, the points in time at which retrieval sets are engaged and, presumably, can influence subsequent retrieval operations, are likely to be influenced by the task that participants are switching from (Herron & Wilding, 2006b).

In another study in which preparatory task-related processes were investigated, the preparatory period (Cue-to-Stimulus Interval, CSI) was lengthened from 2000ms to 4000ms for most of the trials (Herron & Wilding, 2006a). The differences between the activity whilst preparing for an episodic task (location judgments) and a semantic task (movement judgments) were again largest over right-anterior sites (800-4000ms). Consistent with the findings from the previous studies where there were switches between an episodic and non-episodic task (Herron & Wilding, 2004; Morcom & Rugg, 2002) the divergence between neural activities was principally on stay trials. It was concluded that restricting the time available to prepare was not the main factor determining the stay trial onset of this putative preparatory index of retrieval mode. In addition, the activity was attenuated following the preparatory cue on the third trial of the same task (stay+1 trials). This finding further suggests that this ERP index reflects the initial adoption and/or configuration of the retrieval set, rather than the maintenance of it (Herron & Wilding, 2006a).

Moreover, in this study the accuracy of location judgments increased over the number of trials on which participants completed the same task, whereas the accuracy of identifying whether things had been seen before or not (old/new discrimination) remained constant. This evidence suggests that the successful adoption of retrieval mode can influence the availability of recollection of details for task judgments (source memory) under certain circumstances. To explain why switch costs in accuracy were not identifiable in their experiment, Morcom and Rugg (2002) suggested that for a high proportion of old/new trials familiarity was available as the basis for test judgments, and that the availability of this process was not affected by the switching manipulation. This might also explain the absence of an old/new discrimination switch cost in the study reported by Herron and Wilding (2006a).

Wilckens, Tremel, Wolk, and Wheeler (2011) sought to further investigate the premise that recollection relies on the adoption of a retrieval task-set, and that retrieval based on familiarity does not. Wilckens et al. (2011) investigated preparatory as well as item-related effects (the specific item-related effects of interest were the ERP components associated with either familiarity or recollection). Following Herron and Wilding (2006a), Wilckens et al. (2011) also included stay+1 trials.

For the preparatory phase, Wilckens et al. (2011) observed that activity was more positive-going for the semantic than episodic retrieval task on stay trials and largest at anterior-central sites. In this study pictures were used instead of words, and this was proposed as potentially accountable for the reversed polarity in the divergence between these tasks. Across experiment comparisons are also difficult because a global vertex reference point was used, instead of the average of the signal at the two mastoids (employed in all of the studies described above).

Like Herron and Wilding (2006a), Wilckens et al. (2011) also identified behavioural improvements in source memory, and not old/new discrimination, across trial types (although the source memory improvement was for a separate experiment to that in which the ERP and old/new discrimination measures were recorded). In the experiment in which ERP measures were recorded, the left-parietal ERP old/new effect was larger on repeat trials (stay and stay+1 trials collapsed), (this effect has been associated with recollection, see Friedman & Johnson, 2000; and Wilding & Sharpe, 2003, for reviews of old/new effects). The old/new ERP effect that has been associated with familiarity (the mid-frontal effect evident across lateral superior electrodes bilaterally, see Rugg & Curran, 2007) was not modulated by switching. By linking the ERP correlates in one experiment to the behavioural performance in another, the authors concluded that adopting a retrieval task-set impacts on recollection but not familiarity. In addition, Evans, Herron, and Wilding (2012) also reported larger left-parietal ERP old/new effects on stay+1 than on stay and switch trials, and reaction times for accurate context judgments were fastest on stay+1 trials.

1.4.2.3 Item ERPs for correct rejections: additional insights from indices of retrieval orientations

Also noting the potential benefits that are conferred by adopting a retrieval set, Wilding and Nobre (2001) reported slower reaction times in an experiment where participants had to frequently switch between two episodic memory tasks, compared to another experiment where the memory tasks were completed in separate blocks. Furthermore, Wilding and Nobre (2001) observed differences between the ERPs associated with correct rejections only in the blocked experiment design.

Focusing on contrasts between neural activity associated with correct rejections is another approach that has been employed in order to investigate retrieval orientations. The basic rationale is that contrasting ERPs elicited by correctly identified new items under different episodic task demands can reveal processes that form part of a retrieval attempt, and will not be confounded with indices of retrieval success (Ranganath & Paller, 1999; Rugg & Wilding, 2000; Wilding, 1999). Presumably differences between the ERPs elicited by correct rejections are the consequence of having adopted orientations successfully (Rugg & Wilding, 2000).

Robb and Rugg (2002) demonstrated that ERPs elicited by correct rejections differed markedly according to study material. When words rather than pictures were the sought-for material the ERPs were relatively more positive-going. This is consistent with other studies demonstrating differences between correctly classified new item ERPs as a function of the specific episodic retrieval goal (M. K. Johnson, Kounios, & Nolde, 1997; Ranganath & Paller, 1999; Rugg, Allan, & Birch, 2000; Wilding, 1999). In keeping with other accounts, Robb and Rugg (2002) proposed that the differences obtained demonstrated the existence of separate retrieval orientations biasing processing in favour of retrieval of either one kind of content or another (in their case, words or pictures). Additionally, in this experiment there was a 'retrieval effort' manipulation. This was constructed via two different levels of difficulty: by using a short or long study-test delay, and short or long study list lengths. In comparison to the effect of study material, the difficulty manipulation produced a modest statistically significant effect which was small and short-lived. In addition, it did not overlap with the material effect. The authors concluded that replication of this difficulty effect was necessary before any substantial conclusions could be drawn, but, in any case, that it was clear that the material effects (separable retrieval orientations) could not be reduced to difficulty effects alone (Robb & Rugg, 2002; and for other relevant comments see Rugg & Wilding, 2000).

Werkle-Bergner et al. (2005) also investigated differences between ERPs elicited by correct rejections under different task demands. Like Wilding and Nobre (2001), their aim was to examine whether maintaining a retrieval orientation and flexibly adapting to different retrieval demands are dissociable (Werkle-Bergner et al., 2005). However, in Werkle-Bergner et al. (2005) the experimental design was within subjects, enabling a more direct comparison of continuous and alternating task performance. Participants performed four continuous task

blocks, and intermixed between these pure blocks were four blocks where they were required to alternate between two tasks every second trial. Wilding and Nobre (2001) were able to assess ERPs when retrieval demands varied trial-by-trial, however, they were not able to investigate whether such processing emerges only after one trial of a particular task (on a switch trial) or whether multiple trials are necessary (non-switch trials). Werkle-Bergner et al. (2005) planned such comparisons in advance.

Differences emerged between the ERPs elicited by new items on switch and stay trials at right-frontal recording sites (switch > stay and continuous), and it was proposed this reflected selective involvement when a new retrieval orientation had to be activated (Werkle-Bergner et al., 2005). Importantly, however, and in accordance with the findings of Wilding and Nobre (2001), there were indices of separate task-specific retrieval orientations (as indexed by differences between ERPs elicited by correct rejections) in the continuous pure blocks only. The absence of any task-specific divergences in the alternating blocks indicated, again, that adoption of task specific retrieval sets takes a number of successive retrieval trials to emerge in the electrical record (Werkle-Bergner et al., 2005).

J. D. Johnson and Rugg (2006) also analysed preparatory task-related ERPs and item-related ERPs elicited by correct rejections. They too investigated the influence of consistent versus inconsistent retrieval demands (Herron & Wilding, 2006b; Werkle-Bergner et al., 2005; Wilding & Nobre, 2001), however, they employed the exclusion procedure (Jacoby, 1991). Participants were required to accept items corresponding to one class of study material, known as 'targets'. They were required to reject both new items and items from the alternative study material class, known as 'non-targets'. What constituted the 'target' material either varied unpredictably from trial-to-trial, or remained constant throughout a block. J. D. Johnson and Rugg (2006) replicated the results of Herron and Wilding (2006b). They identified differences between the preparatory task cue ERPs for the mixed trial condition only, and ERPs elicited by new items for each task (each targeted study material type) differed during the blocked condition. These outcomes are again consistent with the claim that it takes a number of successive trials to fully adopt a task specific set (Herron & Wilding, 2006b; Werkle-Bergner et al., 2005).

The outcomes in these studies have provided evidence that is relevant to the question of when people are able to prepare for episodic retrieval. Additional important data were reported by Bridger, Herron, Elward, and Wilding (2009), who showed that the extent to which orientations are adopted (as indexed by the size of the differences between ERPs elicited by correct rejections) predicts the accuracy of memory judgments. These data were the first of this kind to indicate this link (for other relevant observations see Rosburg, Mecklinger, & Johansson, 2011; and Rugg et al., 2000).

1.4.2.4   Summary

The combination of task-switching designs and EEG/ERP measures has enabled more precise delineation of the operation and characteristics of retrieval processes than the haemodynamic imaging investigations permitted. Identifying ERP markers of distinct processing stages over time, and investigating the behavioural benefits that are conferred when these indices are observed, has provided a platform for understanding how people selectively retrieve relevant memory traces.

The starting point for the empirical work in this thesis was the observation that the evidence-base for the existence of a common signal that reflects processes linked to retrieval mode has at least one weakness. This weakness stems from the observation that in the studies in which contrasts between preparation for episodic and semantic retrieval have been made the content of what is to be retrieved has also differed. This raises the possibility that differences ascribed to preparing for a certain kind of retrieval are in fact reflective of preparing for different kinds of contents. The background and rationale for work in this thesis is expanded upon at the start of **Chapter 3**, following an account of the methods employed in this thesis.

2. CHAPTER 2. General methods

The following chapter includes a description of electrogenesis and related issues, as well as an overview of the general methods and analysis procedures that were employed in the experiments reported in this thesis.

2.1 The electrogenesis of event-related potentials

ElectroEncephaloGraphy (EEG) is a technique used to directly record the electrical activity of the brain (Luck, 2005). EEG can be recorded intra-cranially. With human participants, however, it is more commonly recorded non-invasively from the outer surface of the scalp (Luck, 2005). Electrical Event-Related Potentials (ERPs) are voltage fluctuations in the electroencephalogram signal that correspond with a physical (exogenous) or mental (endogenous) occurrence (Picton et al., 2000). These ERPs can be extracted from the continuous EEG by filtering techniques and signal averaging (Luck, 2005; Picton et al., 2000).

2.1.1 Research opportunities afforded by the EEG/ERP technique

The millisecond temporal resolution of the EEG technique enables the recording of real-time changes in neural activity (Donaldson, Allan, & Wilding, 2002; Friedman & Johnson, 2000; Rugg, 1995). However, in cognitive studies, the activity elicited on any single trial is not often analysed directly (Donaldson et al., 2002). In the most common approach, the activity elicited on many trials, time-locked to certain stimulus types or 'events of interest', is averaged together. As a result random noise will decrease as a function of the square root of the number of trials in the average (Donaldson et al., 2002; Luck, 2005).

ERP measures can complement behavioural measures, and act as a useful tool for constraining and generating theories of cognitive processing. ERPs recorded from the human scalp have contributed towards the understanding of how healthy human brains process information, enabling subsequent investigations of how this processing may be affected in neurological or psychiatric disorders (Picton et al., 2000).

Generally, inferences are made about cognition by exploring the time course of ERPs, changes in magnitudes, and changes in their distribution across the scalp. The following section contains a description of how the brain generates these ERPs, which is important for an appreciation of potential issues whilst acquiring and interpreting this neural measure (Luck, 2005).

2.1.2 *Neuronal electrogenesis*

Scalp-surface recorded EEG reflects volume conduction of the electrical activity from large populations of neurons within the brain, as well as capacitor conduction through structures such as: the dura membrane, skull, scalp and electrode (Jackson & Bolger, 2014). This electrical activity originates from the changing flow of positive and negative ions across the cell membranes of individual neurons (when neurotransmitters bind to receptors); such changes are referred to as post-synaptic potentials (Luck, 2005). However, this signal from single neurons is not large enough to be detected at the scalp surface. Thus, what is recorded is the summation of synchronous activity from populations of cells (Jackson & Bolger, 2014).

In addition, in order for a detectable signal at the scalp surface, the neurons must be arranged in a parallel fashion. In other configurations, the positive and negative ends of the diploe associated with each neuron cancel each other out (Jackson & Bolger, 2014). The dipole is the term used to describe the situation whereby two regions of positive (the source) and negative (the sink) charge are separated along a neuron (Jackson & Bolger, 2014) (**Figure 1**).
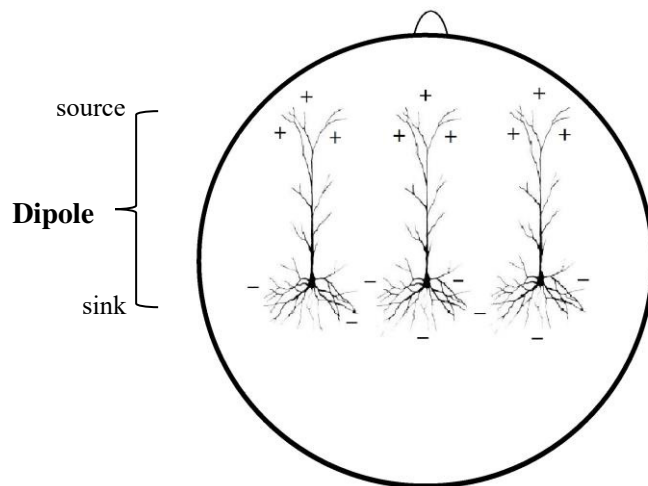
Figure 1. Dipole resulting from the synchronous activity of multiple neurons in parallel. (Pyramidal cell drawing from: http://education.jhu.edu/sebin/w/o/Brain.gif).

The recording of electrical activity is presumed to primarily reflect mass coherent action of pyramidal cells located in the cerebral cortex (Luck, 2005). These are the primary input-output cells of the cortex, and are aligned perpendicular to its surface (Luck, 2005). Parallel-aligned neurons close to the cortical surface are detectable at the scalp, however, non-laminar orientated neurons and interneurons within the cortex are thought to generate little or no detectable activity (Luck, 2005). Thus, activity that is produced from cell populations deeper than the cortical surface, from cells that are not orientated in parallel, and not firing synchronously, is not captured by EEG scalp recording techniques. As a consequence, the recording on the scalp surface reflects only a non-uniform fraction of overall brain activity.

2.1.3   Weaknesses of the EEG/ERP technique

It is important to note that, because the orientation of dipoles determines what is recorded at the scalp surface, inferences cannot be made about whether positive- or negative-going activity reflects excitation or inhibition. This is because each dipole reflects both positive and negative charge in opposite separable poles. It depends on the particular orientation of the dipole as to which direction this charge flows (Jackson & Bolger, 2014), and, in addition, the particular orientation one is recording from.

As the recorded signal reflects the summation of charged ions from large populations of cells orientated in a particular direction, precise source localisation of specific signals is problematic. Each dipole propagates ion flow in opposing directions, thus, each dipole influences the summated signal at a range of scalp surface sites. It is a common misconception to assume that the signal at the scalp surface reflects dipoles immediately below it (Jackson & Bolger, 2014).

Volume conduction is the summation of potentials and the passage of ion charge in a 'wave' through extracellular space. Moreover, aside from the physical arrangement of neurons, varieties in tissue density and its inherent electrical properties, as well as dipole size, contribute toward the direction and strength of ion flow (Jackson & Bolger, 2014). In addition, there are other physically insulating barriers that contribute to the difficulties in determining the source of the signal. These include poor conductors such as the dura, skull layers, and the scalp, as well as myelin-coated nerve tracts within the brain itself (Jackson & Bolger, 2014). These structures effectively form layers of capacitors whereby the signal is propagated via their insulating properties. The insulating layers ensure that the charges do not mix and cancel each other out. Instead, charge builds up at the inner edge of the layer, resulting in the accumulation of ions of the opposite charge on its outer edge (Jackson & Bolger, 2014). Fissures and foramina in the skull also contribute towards signal smearing, and towards the difficulty in determining source localisation of the recorded EEG signal. Thus, signal may flow more easily via conducting pathways, and does not necessarily stem from immediately underlying brain regions. A variety of barriers may either block or aid the ease of ion flow, and signal propagation (Jackson & Bolger, 2014).

The task of working backward from the surface voltage pattern to determine which sources within the brain produced the signal is known as the inverse problem (Jackson & Bolger, 2014). As there are an infinite number of configurations of the sources that may generate a particular pattern of voltage measured at the scalp, there is no unique solution to the inverse problem (Jackson & Bolger, 2014). However, attempts at source localisation of EEG signals have been made via methods of approximation such as modeling techniques with reductionist approaches and assumptions that simplify the complexity of circuits within the brain, as well as the inter- and intra-individual variability in brain structure (Jackson & Bolger, 2014).

41

2.2   Acquisition

2.2.1   Electrode placement

Scalp-surface EEG is typically recorded from electrodes arranged on the head according to standardised systems for electrode placement (Picton et al., 2000). The most commonly used is the international 10-20 system (Jasper, 1958). This system bases electrode placement on 10 and 20 percent increments of the distance between the nasion and the inion (see **Figures 2 & 3**). Modified combinatorial nomenclature were later introduced by Sharbrough et al. (1991), including additional halfway sites between the traditional 10-20 percentage breakdown. This modification accommodated denser arrays of electrodes, and T3 and T4 were relabeled T7 and T8, respectively. The electrode montage used in the EEG experiments throughout this thesis comprised the 25 scalp sites as positioned in **Figure 2**: including midline (Fz, Cz, Pz) and left/right hemisphere locations for the fronto-polar (Fp1/Fp2), frontal (F7/F8, F5/F6, F3/F4), central (T7/T8, C5/C6, C3/C4), posterior (P7/P8, P5/P6, P3/P4), and occipital (O1/O2) sites.
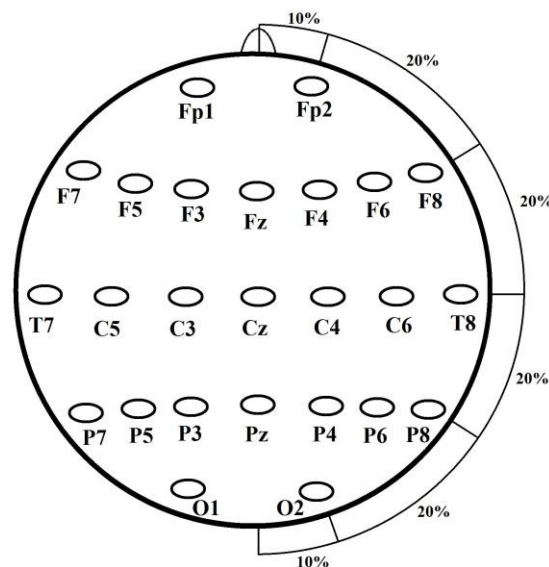
Figure 2. Schematic of the scalp recording locations with standardised naming conventions from the modified 10-20 system (Sharbrough et al., 1991). This nomenclature includes the fronto-polar (Fp), frontal (F), central (C), temporal (T), parietal (P), and occipital (O) scalp sites. Smaller numbers are located adjacent to the midline, with odd numbers to the left, and even numbers to the right.

Recording from multiple electrode sites means that multiple processes can be disentangled via their distribution across the scalp. In addition, recording from numerous scalp sites allows the contribution of artifactual potentials to be identified, and enables different components in the ERP to be captured at optimal locations (Picton et al., 2000).

For scalp-surface recording, the electrodes are commonly attached using plastic housings within an elastic cap. Cap size is selected from a range, based on the circumference of the participant's head, and is used to position the electrodes according to the 10-20 percent increments. The placement of the cap is determined using fiducial points (including the nasion, inion, and preauricular areas, see **Figure 3**) (Picton et al., 2000). The Zenith electrode (Cz) is positioned halfway between the nasion and inion, and halfway between each preauricular area. Thus, electrode placement according to this system allows for variation in head shape and size to be compensated for to some extent. The caps used to collect the data for this thesis were from the EASYCAP GmbH range, Brain Products (UK) Ltd (http://www.brainproducts.com/productdetails.php?id=20), and single silver/silver chloride (Ag/AgCl) electrodes were used.
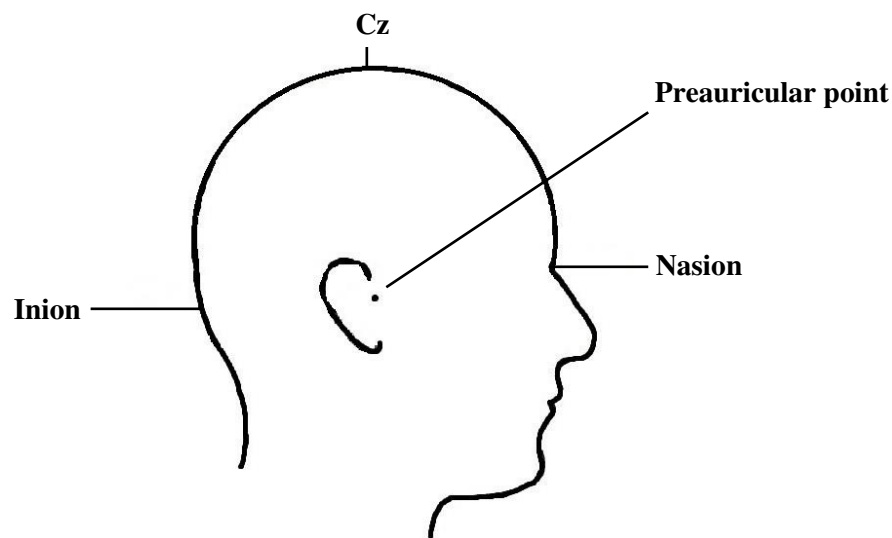
Figure 3. Schematic of the fiducial points used for electrode positioning (nasion, inion, and preauricular areas).

Additional electrodes were also placed on the face of the participant in order to monitor and record eye movements throughout the recording session. These were positioned above and below the right eye, and on the outside of each of the eyes. Electrodes were also placed on the mastoid processes (behind each ear of the participant) for re-referencing after data collection (see **Figure 4**). These electrodes were attached using double-sided adhesive collars.
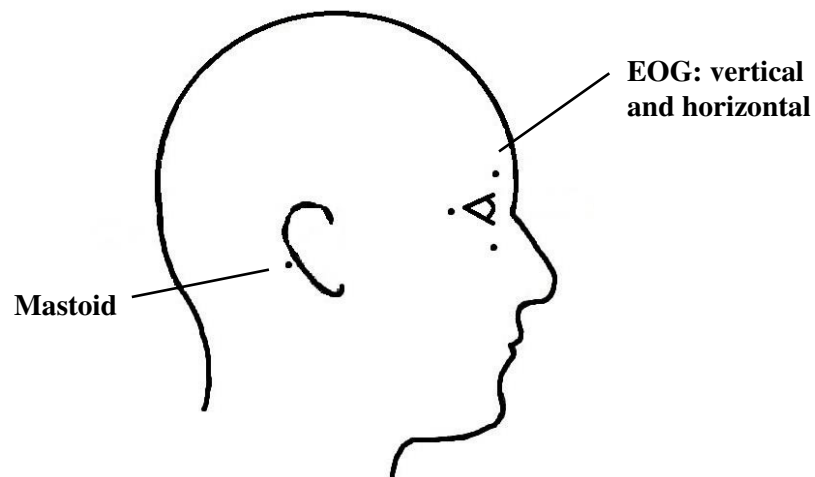
Figure 4. Positioning of Electro-OculoGram (EOG) bipolar electrodes (vertical and horizontal pairs), and re-referencing sites: the mastoid processes.

2.2.2   Noise and artefacts

The electrodes record electrical activity from the brain, as well as other sources of activity which can generically be referred to as 'noise'. Forms of external noise can be limited during online recording. External noise may constitute electrical interference from other electrical devices, such as other equipment or the power supply in buildings: in the floor, walls, ceilings and lights (Jackson & Bolger, 2014; Keil et al., 2014). The signal-to-noise ratio (S:N) refers to the measure of how much signal is reflected in the recording, relative to unwanted noise sources (Jackson & Bolger, 2014). An amplifier is used during acquisition to increase the signal of the voltage change measured. This is so that any noise introduced later in the circuit (perhaps as the activity travels along wires to the acquisition computer) is small relative to the earlier amplified signal (Jackson & Bolger, 2014).

When participants blink the muscle movement is propagated across the scalp surface, and adds noise to the recording at the scalp surface electrodes. This is a form of internal noise that is monitored via the EOG electrodes, and can be subtracted during later pre-processing stages. Other internal noise sources include sweat potentials, muscle tensions, or brain activity that is not related to the research question under investigation (Jackson & Bolger, 2014; Keil et al., 2014).

Artefact rejection is routinely carried out during pre-processing to remove noise related to the subject, as well as remaining noise from the external recording environment and equipment (Keil et al., 2014). Such artefacts contaminate the recording, and removal is necessary in order to obtain a clean signal for analysis. For the EEG experiments reported in this thesis, removal of artefacts was carried out using a combination of visual inspection and automatic detection algorithms. The algorithms for automatic detection of artefacts included: setting the minimum and maximum allowed amplitude (-100, +100μV) and absolute difference (200μV/200ms), setting the gradient voltage step per sampling point (75μV/ms), and setting detection for low activity levels over time (0.5μV/50ms).

2.2.3   Impedance

2.2.3.1   Input impedance of the amplifier

The amplifier increases the S:N ratio via a set input impedance which determines how well the system can tolerate weaker signals (Jackson & Bolger, 2014). The input impedance is set as large (>1TΩ), and, effectively, the amplifier acts as a voltmeter applying a voltage drop across it. With a large resistance, more of the voltage is measured (including high recordings of both the signal and the noise). The noise experiences less resistance, depending on where it comes from and what stage it influences the circuit, as it travels less distance along the wire before it reaches the amplifier. Thus, without the large input impedance, the signal would be attenuated (proportionally) more than the noise. However, with a large input impedance set on the amplifier, the ratio of S:N is closer together. This is because the amplifier measures a large amount of signal amongst a large amount of noise (Jackson & Bolger, 2014).

45

The amplifier used during data collection for this thesis was the QuickAmp, from Brain Products (UK) Ltd. This amplifier also has active shielding technology. This means that noise picked up by the shielding along the wire cable for each channel is fed back to the amplifier where it is actively suppressed (Brain Products Brain Vision (UK) Ltd, 2014).

2.2.3.2 Impedance at the scalp surface

Another form of resistance is the impedance between the scalp and the electrode. In this instance, small impedance is desirable. This is because here the signal is passed to the electrode, rather than measured. With smaller impedance the signal is passed more easily, because with a smaller voltage drop more signal passes through the circuit (Jackson & Bolger, 2014).

Electrode gel is necessary in order to provide the conductive path from the scalp to the electrode (Jackson & Bolger, 2014). Within the brain the cerebral spinal fluid and various other substances are very good conductors. These are, however, separated from the electrode by the series of poor conductors that comprise the capacitor layers (Jackson & Bolger, 2014). In order to saturate the space that remains between the scalp and the electrode imbedded in the elasticated cap, highly conductive electrode gel is used to fill in the air pockets and act as another capacitor layer (Jackson & Bolger, 2014). The impedance is then measured at each electrode site, and commonly adjusted prior to data collection to insure that it is below 5KΩ (Picton et al., 2000). The impedance at the scalp surface should be less than the input impedance of the amplifier (typically >1TΩ) by a factor of at least 100 (Picton et al., 2000). Lowering the impedance at the scalp surface typically requires abrasion of the skin to remove any dead skin cells, and is carried out using a cotton swab during the application of electrolyte gel. In addition, the hair of the participant is parted to allow the electrolyte direct contact with the scalp surface. These approaches were used for the EEG experiments in this thesis.

2.2.4 Referencing

During data acquisition for the EEG experiments in this thesis, the voltage at each site was recorded relative to a single average reference. This is calculated as the sum of the activity in all recording channels divided by the number of channels plus one (Picton et al., 2000).

A different reference is frequently used for online recording and offline analysis (Keil et al., 2014). The data were re-referenced after acquisition (during pre-processing) to the average of the signal at the two mastoid processes. Re-referencing allows unwanted activity that is related to all electrode sites to be removed from the continuous electroencephalogram recording. The reference that is selected impacts on the overall morphology of the ERPs, thus, it is important that reference selection is consistent across experiments where comparisons are drawn. The most common reference site is the average of the signal at the two mastoid processes (Rugg & Coles, 1995). The linked mastoids act as a suitable reference as hemispheric differences are ameliorated, recording from these sites is comfortable for the participant, and they are not influenced substantively by brain activity (Nunez & Srinivasan, 1981). This reference is common in published experiments investigating electrophysiological indices of preparation for episodic memory retrieval.

2.2.5   Grounding

Participants were also connected to a ground on the amplifier system via an electrode placed at FCz. Grounding is carried out universally to reduce artifact and leakage currents that flow through the participants (Pivik et al., 1993).

2.2.6   Analog-to-Digital (A/D) conversion and sampling rate

It is necessary to convert the analog signal into digital format to process the data (Pivik et al., 1993). The A/D units are transferred into physical units, and the data are subsequently presented in microvolts (μV) (Keil et al., 2014). The rate of A/D conversion is known as the sampling rate. The sampling rate is guided by the Nyquist theorem, which states that signal frequencies which are equal to or greater than half of the Nyquist frequency (the sampling frequency) will be distorted (Keil et al., 2014). Thus, the extent to which the digital format represents the analog signal depends on the sampling rate (Pivik et al., 1993). Distortion of the signal in this manner is known as aliasing, whereby low frequency components are produced that cannot be separated from those of the true signal (Dumermuth, Ferber, Herrmann, Hinrichs, & Künkel, 1987). The sampling rate must be at least twice the highest frequency present in the signal to avoid this distortion. During data acquisition for the EEG experiments included in this thesis the sampling rate was 250Hz. This is the number of samples taken per second, and thus, a sample was taken every 4ms.

2.3   Additional data pre-processing steps

2.3.1   Filtering

In order to eliminate noise with spectral content that does not overlap with the frequency content of the signal of interest, such as mains noise and drift, filtering of the data is typically employed. For the EEG experiments contained in this thesis, low and high pass filters (0.03-40Hz, 24dB/oct) were applied to the continuous data.

2.3.2   Ocular correction

The remaining data were then used to run an ocular correction algorithm. Both vertical blinks and horizontal saccades influence the EEG recording, as muscular movement of the eyelid and lateral movements of the eyeball propagate current over the head. As generally recommended, participants were instructed to minimise eye movements throughout the task (Luck, 2005; Picton et al., 2000). However, as is also standard protocol during the pre-processing of EEG data, an algorithm was used to correct for any eye movements that occurred during the task (Keil et al., 2014). This process estimates the relative contribution of eye movements to the continuous EEG signal, and removes it from the recording accordingly (Gratton, Coles, & Donchin, 1983). In addition, trials containing any remaining EOG artefact were later rejected during visual inspection of the EEG waveforms for each channel associated with each trial.

2.3.3   Differential segmentation of ERPs

The remaining data were then separated into the conditions of interest via grouping according to specific stimulus events and in some cases participant responses.

2.3.4   Baseline correction

For each trial an additional pre-stimulus period was included in the epoch. This is known as the baseline period, and the mean activity during this time acts as a zero value for any changes in voltage to be quantified against (Keil et al., 2014). During the pre-processing of the data included in this thesis a baseline period of 200ms was used, as suggested by Luck (2005).

2.3.5 Averaged ERPs

Following this series of pre-processing steps, trials within each differentially segmented group of ERPs were averaged together. In keeping with the approach adopted in previous studies of preparatory ERP modulations (e.g. Herron & Wilding, 2006a), in each EEG experiment in this thesis these averages included at minimum 16 artefact-free trials per condition per participant. This was to ensure that any signal related to each event of interest was likely to be distinguishable from other noise or separate processes that are not part of the current investigation.

2.4 Presentation of data

Topographic scalp maps, displaying differences in voltage over space, were constructed via projection onto a spherical model of the scalp. These are known as spherical spline interpolations (Picton et al., 2000). Interpolation is the mathematical technique used to estimate the data between the electrode locations (Keil et al., 2014). Thus, it is important to note that generally over 99% of the pixels on a scalp map are an interpolation based on the small amount of the scalp surface from which data are obtained (Pivik et al., 1993). The actual recorded data points are obtained from the electrode recording locations in the standard 10-20 layout, and grand-averaged ERPs (across participant averages) from these points were plotted graphically as waveforms. In these plots, changes in voltage are displayed over time at the different electrode locations, and the ERPs for different conditions of interest are overlaid for comparisons (Donaldson et al., 2002; Picton et al., 2000).

2.5 Analysis approaches

The ERPs elicited in different experimental conditions are typically analysed by contrasting time courses, amplitudes, and scalp distributions (Donaldson et al., 2002). This allows for investigation of both quantitative and qualitative changes in neural activity (Donaldson et al., 2002). Quantitative changes are revealed via comparison of amplitude differences between the ERP waveforms, whereas qualitative changes are inferred from differences between the scalp topographies of the waveforms. Across conditions, differences between scalp distributions are frequently used to argue that not entirely the same brain regions, and thus cognitive processes, are engaged (Donaldson et al., 2002).

According to traditional frequentist statistical methods, throughout this thesis the *p* value < 0.05 was adopted for all orthodox statistical analyses. Marginal trends are also referred to, and labelled appropriately, when *p* > 0.05 and (commonly) < 0.1. Furthermore, planned comparisons as well as subsequent exploratory analyses are referred to as such.

During EEG data acquisition, simultaneous recordings are collected from multiple channels over time. Thus, a great number of observations are recorded. Numerous combinations of different timescales (epochs), electrode sites, and frequency bands could in principle be compared (Pivik et al., 1993). This may lead to significant effects in the form of false positives (Kilner, 2013), purely due to the sheer number of comparisons made (Keil et al., 2014). When multiple tests are conducted the probability of making a type one error and incorrectly concluding that there is a significant effect from the manipulation increases (known as alpha inflation) as one can choose measures that (inadvertently) take advantage of the noise in the data (Keil et al., 2014; Simmons, Nelson, & Simonsohn, 2011).

Multiple corrections exist in order to cope with the alpha inflation of *p* values, and, thus, to control for false positives. However, in EEG research typically, and appropriately, the multi-dimensional dataset is reduced by collapsing across various measurements according to the specification of prior research (Keil et al., 2014; Kilner, 2013). This prior specification of the parameters of interest from independent datasets is necessary in order to constrain the analysis approach selected. Throughout this thesis, the data were reduced by computing mean amplitude measures for the ERP activity associated with conditions of interest. In addition, the epochs and the sites submitted to analysis were specified *a priori*, guided by previous literature demonstrating the effects of interest.

In order to reduce the number of variables, further condensing methods were also utilised (Pivik et al., 1993). Where appropriate, the data were collapsed across sites in the electrode analysis montage. For instance, factors submitted to each global repeated measure ANalysis Of VAriance (ANOVA) included: hemisphere, and the anterior-posterior dimension. Throughout this thesis, this condensing approach has also been conducted in accordance with the previous literature demonstrating the effects of interest. This allows for

inferences to be made regarding the lateralisation and location of effects in the anterior-posterior plane, as per hypotheses regarding divergences across conditions.

For each repeated measures ANOVA the Greenhouse-Geisser correction for non-sphericity was applied where necessary (Greenhouse & Geisser, 1959). Repeated measures ANOVAs assume sphericity, whereby the variances of the difference scores between all pairs of the within-subjects levels are equal. With EEG datasets there is an array of data points that are highly related, thus, correction for non-sphericity is typically necessary.

The sphericity index (the Box index; Box, 1954a, 1954b), is denoted ε (epsilon) and varies between zero and one (Abdi, 2010; Picton et al., 2000). When a co-variation matrix has the property of sphericity the value of the epsilon is one, and the more severe the violation of sphericity the smaller the value of the epsilon (Abdi, 2010; Picton et al., 2000). The Greenhouse-Geisser correction computes the epsilon and adjusts for a violation accordingly (Greenhouse & Geisser, 1959). The degrees of freedom for all effects involved in the within subjects analysis are multiplied by the epsilon factor. Therefore, the degrees of freedom are reduced, and this makes the *F*-test more conservative as a larger effect is required to reject a critical value with fewer degrees of freedom. Consequently, violation of the sphericity assumption results in a loss of statistical power because of the need to adjust the probability values to correct for the violation. Throughout this thesis the epsilon-corrected degrees of freedom are reported where necessary.

2.5.2 *Power analyses and effect size computation*

With the rise in the importance of the replication of research findings, increasing attention is being paid to measures of effect size as well as statistical power analyses (Button et al., 2013; Keil et al., 2014; Lakens, 2013). The effect size measurement known as eta-squared (η^2) is effective for comparing the sizes of effects within a single study. This is because within an experiment this effect size is interpreted in relation to the total variance possible. However, η^2 should not be compared across studies, as the total variability within a study depends on its design, and the number of variables that are manipulated (Lakens, 2013). The effect size known as partial eta-squared (η_p^2) was recommended by Keppel (1991), in order to improve the relationship between effect sizes across different studies. This measure expresses the sum of squares of the effect whilst taking into account the sums of

51

squares of the error associated with it, rather than the total variability within the single experiment (Lakens, 2013).

Cohen's *d* (Cohen, 1988) is frequently used to represent the standardised mean difference of an effect. For example, for meta-analytic comparison across similar studies with different designs or analysis strategies (different electrode sites/clusters/epochs etc.). However, Cohen's *d* benchmarks for effect size interpretation (**Table 1**) were not originally envisaged to be used for effect sizes that take into account correlated measures in within-subjects designs (Lakens, 2013). Thus, for within-subjects designs, the standardised mean difference effect size is referred to as Cohen's d_z (Lakens, 2013). This effect size measure takes the correlation between measurements into account, and the standardiser is the standard deviation of the difference scores (Lakens, 2013).

The effect size measures reported throughout this thesis are Cohen's d_z for within-subjects main effects between two factors, and partial eta^2 (η_p^2) for main effects between three factors and interactions. These specific measures are the required input for power analysis computations using G*Power3.1.7 software (for interactions the programme converts η_p^2 to Cohen's *f*: see **Table 1** for interpretation of Cohen's *f*). Cohen's d_z was calculated using the formula: $d_z = t/ \sqrt{N}$ (Lakens, 2013; Rosenthal, 1991). Partial eta^2 was computed using SPSS, which reflects the relationship between η_p^2 and *F*; $\eta_p^2 = F$ x $\text{df}_{\text{effect}}$ / *F* x $\text{df}_{\text{effect}}$ + df_{error}. This equation illustrates how η_p^2 can be used in *a priori* power analyses to estimate the sample size required to achieve a certain significance level in an *F* test (Faul, Erdfelder, Lang, & Buchner, 2007; Lakens, 2013).

2.5.2.1 *Effect size: measurement interpretations*

Table 1 shows the effect size measurements and their respective interpretive benchmarks. Cohen (1988) provided these benchmarks to define small, medium, and large η^2 effects, so that comparisons could be made between unrestricted populations (Olejnik & Algina, 2003). However, according to Olejnik and Algina (2003), as Lakens (2013, p. 7) states: "using these benchmarks when interpreting the η_p^2 effect size in designs that include covariates or repeated measures is not consistent with the considerations upon which the benchmarks were based". In any case, G*Power3.1.7 converts η_p^2 to Cohen's *f* values in order to compute *a priori* power analyses for interactions, and main effects between three

factors (Faul et al., 2007). The interpretive benchmarks for Cohen's f effect sizes are also included in **Table 1**.

Similarly, Cohen (1988) defined d's as small, medium, and large effects. According to Lakens (2013) $d_s \sim 2 \times t/\sqrt{N}$, and $d_z = t/\sqrt{N}$. However, using Cohen's benchmarks to interpret effect sizes is not best practice, and it is more meaningful to relate the effect size to other similar effects in the literature (Thompson, 2007). Lakens (2013) recommends use of the Common Language (CL) effect size statistic. The CL reflects, intuitively, the probability that an individual participant has a greater value on one measurement than another (McGraw & Wong, 1992). Lakens (2013) includes a supplementary spreadsheet which aids computation of this value for within-subjects main effects. According to Preacher and Kelley (2011) reporting multiple effect sizes can improve understanding of a single effect. Thus, throughout this thesis, the CL is reported alongside Cohen's d_z measures.

Table 1. Effect size measurement interpretations: useful benchmarks (small, medium, large) (Cohen, 1988; Faul et al., 2007).

	Eta-squared (η^2)* (0-1 range)	Partial eta-squared (η_p^2)	Cohen's f	Cohen's d	Cohen's d_z
Small	0.02	-	0.10	$\geq$0.20	Convert to CL
Medium	0.13	-	0.25	$\geq$0.50	Convert to CL
Large	0.26	-	0.40	$\geq$0.80	Convert to CL

* For repeated measures ANOVA

2.5.2.2 A priori power analyses

In frequentist Null Hypothesis Significance Testing (NHST), the data are compared to the null hypothesis. The probability of correctly rejecting the null hypothesis is what is referred to as the *power* of a statistical test (1 - β) (Cohen, 1988; Faul et al., 2007). The symbol β represents the Type II error rate (beta error probability) for falsely retaining the incorrect null hypothesis (Faul et al., 2007). This power of a statistical test depends on the sample size within the study, because the sample value is expected to represent the population value of the effect, and with larger samples this sample value is more reliable (Lakens, 2013).

With low statistical power, significance testing is limited because it is unable to discriminate between the null and alternative hypotheses (Faul et al., 2007). The generally accepted minimum level of statistical power is 0.80 (Cohen, 1988; Lakens, 2013). This is because with power set at 0.80, the probability of a Type II error in relation to a Type I error is reflected as 0.20:0.05. Whereby the probability of concluding there is an effect when there is not is reflected as four times as problematic as concluding there is not an effect when there is, in the given population (Lakens, 2013).

There are different types of power analyses, depending on the output parameters one wants to know. These include: a priori, post-hoc, compromise, sensitivity, and criterion power analyses techniques (Faul et al., 2007). *A priori* power analyses allow researchers to gather an indication of the average sample size required to observe a statistically significant effect, in a replication attempt of an original study (Faul et al., 2007; Lakens, 2013).

For *a priori* power analyses, the average sample size required (N_{req}) is computed for the required power level ($1 - \beta$), the pre-specified significance level (α), and the population effect size to be detected with probability $1 - \beta$ (Cohen, 1988). For the *a priori* power analyses carried out before data collection and reported in this thesis, the estimated sample size (N_{req}) was computed for statistical power set at 0.80 (Cohen, 1988), the pre-specified alpha level (0.05), and the population effect size to be detected in each instance (see **Appendix A**). Calculations were computed using the software G*Power3.1.7 (Faul et al., 2007).

Power analyses provide a point estimate for the minimal sample size (Lakens, 2013). Thus, Lakens (2013) advised that these calculations are taken into account whilst bearing in mind that the true underlying effect size is uncertain. It is inadvisable to focus solely on *a priori* power analysis for determining sample sizes, because the original published effect is probably an overestimation. This is due to issues known as: publication bias (Lane & Dunlap, 1978), subsequent regression to the mean (Bland & Altman, 1994), the Proteus phenomenon following the winner's curse (Button et al., 2013), and the likelihood of false positives in the literature (Button et al., 2013). Thus, throughout this thesis, *a priori* power analyses were taken into account, as well as counterbalancing demands, and the Bayesian Stopping Rule (explained further in the following subsection) (Dienes, 2011; Edwards, Lindman, & Savage, 1963; Verhagen & Wagenmakers, 2014).

2.5.3   Complementary Bayesian analyses

As part of The New Statistics movement, Cumming (2013) has advocated interpretation via estimation. The use of effect sizes and meta-analytical techniques were encouraged, and a shift towards the increased reliance on confidence intervals was strongly promoted. It was further suggested that there is no place for traditional NHST in a new research strategy that facilitates cumulative science. This stance is based on what are seen as the inherent flaws around using dichotomous decision-making centred on the arbitrary cut-off, $p < 0.05$ (Cumming, 2013). This proposal followed debates in the literature regarding a 'crisis of confidence' in experimental psychology and the neurosciences, relating to discussions about false-positive (Ioannidis, 2005; Simmons et al., 2011) and false-negative psychology (Fiedler, Kutzner, & Krueger, 2012).

Conversely, Morey, Rouder, Verhagen, and Wagenmakers (2014) subsequently argued that estimation alone is insufficient, and that proper hypothesis testing methods are crucial. According to Morey et al. (2014), hypothesis testing is the basis of the scientific method, and the nature of science is that it is driven by theories that make clear testable predictions. However, Morey et al. (2014) agreed that any reliance on traditional NHST should be avoided. As a substitute, Morey et al. (2014) promoted the use of Bayesian model comparisons. In addition, in order to support the accumulation of conclusive answers to a variety of questions whilst using a wide range of approaches, Morey et al. (2014) advocated the use of both estimation and Bayesian hypothesis testing side-by-side. Estimation aids pre-theoretical work, as well as theory revision post data collection. However, according to Morey et al. (2014), hypothesis testing is required in order to test the quantitative predictions of theories. Thus, these approaches provide complementary answers to different questions.

Throughout this thesis, in order to investigate whether the results obtained provide quantitative support for the null hypothesis, Bayesian statistics were utilised. Bayesian statistical analysis applies Bayes Theorem and inference, to enable investigation of whether the results are in support of one of two hypotheses (e.g. the null or alternative of the theory proposed) (Dienes, 2011; Edwards et al., 1963; Verhagen & Wagenmakers, 2014).

Providing quantitative support for the null hypothesis cannot be done using traditional NHST (Dienes, 2011; Morey et al., 2014; Verhagen & Wagenmakers, 2014). This is because with orthodox *p* value statistics one can either provide evidence for the existence of an effect, or one can fail to show evidence that an effect exists (Dienes, 2011). Put simply: you can either reject or retain the null hypothesis. With traditional NHST one is investigating whether the results obtained across groups or factors are significantly different, with a 5% probability that the differences are due to chance and a 95% confidence interval. By saying something is 'significant' the researcher is stating that the results obtained were unlikely to have been due to chance alone. However, when the results are not statistically 'significant' from each other, one can only conclude that according to the sampling distribution specified one is unable to reject the null.

More specifically, traditional NHST is based on the initial premise that the null hypothesis is true, and a *p* value is calculated to reflect this assumption. Thus, trying to provide evidence for the null hypothesis using a *p* value is not possible, due to the issue of circularity: because it is attempting to prove it is true based on the assumption that it is true. In summary, traditional hypothesis testing assumes the null is true, and then uses that assumption to *only* allow one to demonstrate evidence that builds a contradiction against that being the case (Dienes, 2011).

Alternatively, with Bayesian analysis one can provide support for one of two hypotheses (Dienes, 2011; Verhagen & Wagenmakers, 2014), where one may be the null. Bayesian analysis compares whether the data is more likely according to one of the two hypotheses. The unit of measurement is the Bayes Factor (BF), and it compares one theory/model against another (theory 1 to theory 2, e.g. the null, H_0, to the alternative, H_1) (Dienes, 2011).

The BF is computed according to Bayes Rule (derived from Bayes mathematical theorem of probabilities) (Dienes, 2011). Bayes Rule expresses the BF, which is the likelihood ratio of the relationship between the prior and posterior probabilities for an effect. The prior probability is the odds of an event (e.g. H_0) in relation to another event (e.g. H_1). The posterior probability is the prior probability related conditionally to a third event. The BF is simply the ratio of the conditional probabilities of the event (observed evidence) given that H_0 or H_1 is the case.

This is calculated according to the equation:

$$\boxed{\textbf{Posterior odds = prior odds x BF}} \quad \text{(Dienes, 2011).}$$

This rule reflects how the judgment of the true hypothesis (H_0 or H_1) is updated following the observation of additional evidence (Verhagen & Wagenmakers, 2014). Thus, the probability estimate for a hypothesis is updated as additional evidence is acquired.

BFs are used throughout this thesis to complement other analysis tools and estimation measures such as: traditional frequentist statistics, effect sizes and confidence intervals. They are useful as they can provide additional supporting statistical evidence to bolster the conclusions drawn (Morey et al., 2014; Wetzels et al., 2011).

BFs quantify the strength of the evidence for each hypothesis and its likelihood (Verhagen & Wagenmakers, 2014). Specifically, BFs greater than one support H_1, whereas BFs less than one support H_0. In addition, the likelihood of the data occurring under one hypothesis in relation to another can also be expressed intuitively as a ratio. For instance, a BF of 20 would indicate that the newly observed data are 20 times (20:1) more likely to have occurred according to H_1 than H_0. A BF of 0.05 would indicate that the newly observed data are $1/0.05 = 20$ times (20:1) more likely to have occurred according to H_0 than H_1 (Verhagen & Wagenmakers, 2014).

Essentially, BFs vary between zero and infinity and the outcome of a BF test is not a reject or accept decision like traditional NHST. BFs function to "grade the decisiveness of the evidence" (Jeffreys, 1961, p. 432), and, thus, there is a continuous scale of values. However, Jeffreys (1961) proposed discrete categories in order to facilitate scientific communication and to allow ease of interpretation when reporting the results of replication tests. Accordingly, values greater than three are taken as substantial evidence for the alternative hypothesis whereas values less than 0.33 indicate substantial evidence for the null hypothesis (Wetzels et al., 2011). The arbitrary cut-off boundaries proposed by Jeffreys (1961) are summarised in **Table 2** (adapted from Table 1 in Wetzels et al., 2011).

Table 2. Discrete categories to aid interpretation of BF Replication Test results.

BF range	Interpretation
>100	Decisive evidence for H_1
30-100	Very strong evidence for H_1
10-30	Strong evidence for H_1
3-10	Substantial evidence for H_1
1-3	Anecdotal evidence for H_1
1	No evidence
0.33-1	Anecdotal evidence for H_0
0.1-0.33	Substantial evidence for H_0
0.033-0.1	Strong evidence for H_0
0.01-0.033	Very strong evidence for H_0
<0.01	Decisive evidence for H_0

For a comparison, the interpretations that follow from the result of p value statistics are summarised in **Table 3** (again adapted from Table 1 in Wetzels et al., 2011).

Table 3. Evidence categories for the interpretation of p values.

p value	Interpretation
<0.001	Decisive evidence against H_0
0.001-0.01	Substantial evidence against H_0
0.01-0.05	Positive evidence against H_0
>0.05	No evidence against H_0

Calculation of the required BF requires consideration of the specific question that one is attempting to address. Specifically, one may be interested in: (i) whether the effect identified previously is present or absent, (ii) whether the effect size obtained is equal to what was obtained previously, (iii) what the effect size supporting the theory is, once taking all of the previous studies into account, or, (iv) whether the effect size obtained in the current study is comparable to what was identified previously. Depending on the question one wishes to answer, there are different BFs that can be computed which take into account different parameters.

- **Computation of the 'default' Bayes Factor: Independent Jeffreys-Zellner-Siow (JZS) Bayes Factor Test** (Dienes, 2011; Rouder, Morey, Speckman, & Province, 2012; Rouder, Speckman, Sun, Morey, & Iverson, 2009)

This test addresses the question: "*Is the effect present or absent in the replication attempt?*" (Verhagen & Wagenmakers, 2014). Despite the intuitive appeal of this test, often this approach does not take into account data from original experiments in construction of the prior (Verhagen & Wagenmakers, 2014). Here, the replication test can be constructed or scaled according to a standardised effect size that the researcher deems important (e.g. small, medium, or large, according to Cohen's *d* benchmarks) (Rouder et al., 2009). It stems from this that the directional prediction for the difference is often not taken into account. An exception is the Dienes (2011) 'default' Bayes Factor Test. In this test, the effect size from the previous literature is taken into account, and it is used as the standard deviation for the population effect (Dienes, 2011).

- **Equality of Effect Sizes Bayes Factor Test** (Bayarri & Mayoral, 2002)

This test addresses the question: "*Does the effect size in the replication attempt equal the effect size in the original study?*" (Verhagen & Wagenmakers, 2014). For this test, the null hypothesis actually reflects a lack of difference between the effect size for the original and replication attempt. Thus, in this instance, support for the null hypothesis corresponds to successful replication.

- **Fixed-effect Meta-analysis Bayes Factor Test** (Rouder & Morey, 2011)

This test addresses the question: "*When pooling all data, is the effect present or absent?*" (Verhagen & Wagenmakers, 2014). This test assumes that all experiments are exchangeable, with the same underlying effect size, which is uncontaminated by publication bias, and that fluctuates only due to sampling variability (Verhagen & Wagenmakers, 2014). Thus, the BF reflects the extent to which the pooled data support the hypothesis for the theory proposed. For example, if one of two effect sizes was low, then the other would have to be substantially higher in order to pull the average up and produce a BF in favour of the effect (Verhagen & Wagenmakers, 2014). Thus, if the effect size in the original experiment is large, but small in the replication attempt, the overall combined BF may still provide evidence in favour of the effect (Verhagen & Wagenmakers, 2014).

- **The New Replication Test** (Verhagen & Wagenmakers, 2014)

This test addresses the question: "*Is the effect from the replication attempt comparable to what was found before, or is it absent?*" (Verhagen & Wagenmakers, 2014). This replication test allows for the prior effect size estimate to constitute an entire distribution of effect sizes from an original study, and not just a point estimate. Thus, this means that the sample size of the original experiment is also taken into account (Verhagen & Wagenmakers, 2014). This test provides a BF and a plot, for an intuitive interpretation of its output.

Throughout this thesis, BFs were calculated and plotted using the R-version of the new replication test (Verhagen & Wagenmakers, 2014). Computation of replication success using the new replication test simply requires the number of participants and the t value for the main effect of interest (including the minus sign if the effect is in the opposite direction), from the original and replication experiment. Thus, this test does not require specification of a prior distribution, eliminating the popular criticism of Bayesian analysis whereby there is too much flexibility in the choice of the prior (Liu & Aitkin, 2008; but see Vanpaemel, 2010).

The new replication test (Verhagen & Wagenmakers, 2014) addresses the final question from the possibilities listed earlier. Specifically, it tests whether the results of the replication attempt are comparable to what was evident before, or not. With this test, even if the replication effect size is larger than the original effect size then the replication test will indicate support for H_1, because even though H_1 will provide a poor account of the data, it provides a better account than the competing H_0 theory (Verhagen & Wagenmakers, 2014).

The test output includes a plot of the BF ratio: for the prior and posterior at the null hypothesis effect size of zero. The posterior from the original experiment is used as the prior for the effect size in the replication test, and is plotted as a dotted line. The posterior distribution after the data from the replication attempt have been taken into account is plotted using a solid line. The ratio of the ordinates at effect size zero provides the result of the replication test (**Figure 5A-C**). The result of the replication test is also given as a numerical BF value, and these can be interpreted according to the discrete categories cited previously (Jeffreys, 1961; Wetzels et al., 2011). When the ordinates are further apart the distributions are less matched and, thus, the test provides support for the null hypothesis as more mass is assigned to the zero effect size for the posterior than for the prior distribution (**Figure 5B**).

At effect size zero, the height of the prior distribution indicates the support for H_0 before the data from the replication attempt are taken into account. The height of the posterior distribution at the effect size zero indicates the support for H_0 after taking into account the replication data (Verhagen & Wagenmakers, 2014). In addition, the less peaked the posterior distribution is, the less extreme the test outcome (the smaller the sample size in the replication, the shorter the posterior peak; **Figure 5C**) (Verhagen & Wagenmakers, 2014).

A.
In support of the alternative hypothesis

$t_{orig} = 2.43$
$t_{rep} = 2.76$
$N_{orig} = 24$
$N_{rep} = 24$

B.
In support of the null hypothesis

$t_{orig} = 2.5$
$t_{rep} = -3.0$
$N_{orig} = 24$
$N_{rep} = 24$

C.
Smaller sample size in replication

$t_{orig} = 2.5$
$t_{rep} = -3.0$
$N_{orig} = 24$
$N_{rep} = 16$
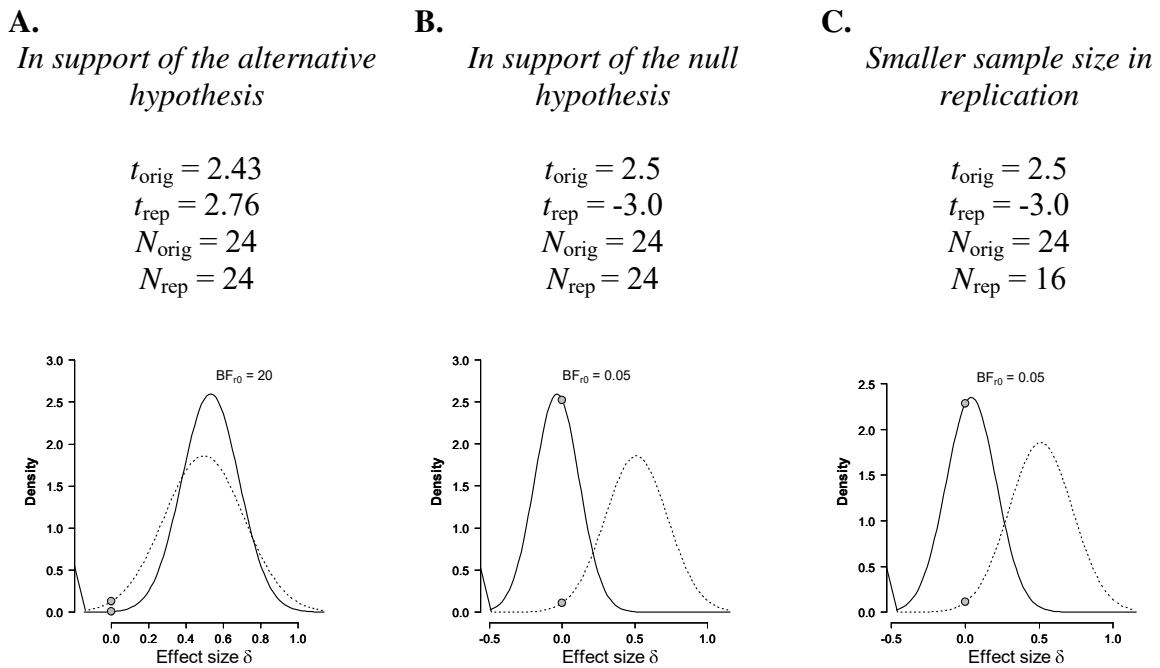
Figure 5. Example data input (*t* values and number of participants for an original study and a replication attempt), output plots of the prior (dotted) and posterior (solid) distributions, and the related ratio Bayes Factors (BFs).

Consistencies between *p* values and BFs have been investigated (Verhagen & Wagenmakers, 2014; Wetzels et al., 2011). *P* values overestimate the evidence against the null hypothesis, and this is because NHST does not take into account the adequacy of the alternative hypothesis (Edwards et al., 1963; Sellke, Bayarri, & Berger, 2001; Verhagen & Wagenmakers, 2014). For instance, if both the null and alternative hypothesis provide equally poor accounts of the data, then in traditional frequentist statistics the *p* value indicates that the null should be rejected. NHST only tests the null hypothesis (reject, or fail to reject). However, intuitively, in this instance, the BF would indicate that the data are ambiguous: providing support for neither the null nor alternative hypothesis (Verhagen & Wagenmakers,

2014). BFs do not assign any special status to one hypothesis over the other, allowing one to quantify evidence in favour of the null or alternative hypothesis (Verhagen & Wagenmakers, 2014).

Whilst using Bayesian analysis, one can monitor the evidence for either hypothesis as the data is collected. Data collection can be terminated whenever there is sufficient evidence in favour of either hypothesis. This is known as the Bayesian Stopping Rule (Dienes, 2011; Edwards et al., 1963; Rouder, 2014; Verhagen & Wagenmakers, 2014). This flexibility is possible whilst computing BFs, because the rules that dictate when data collection should stop are irrelevant for the interpretation of the data (Dienes, 2011; Edwards et al., 1963). Throughout this thesis, the Bayesian Stopping Rule was used in combination with *a priori* power analysis calculations and counterbalancing demands.

Part 1. Electrophysiological indices of preparation for episodic memory retrieval

3. CHAPTER 3. Experiment 1: content matched

3.1 Experiment 1: Abstract

This experiment was designed to delineate the processes engaged while people prepare for memory retrieval. This was achieved by comparing ERPs elicited by two preparatory cues signalling that participants should prepare for different retrieval tasks. One cue signalled preparation for episodic retrieval: identify the location in which an object had been seen in a prior study phase. The other signalled preparation for semantic retrieval: identify the location in which an object is most commonly found. Participants were cued trial-by-trial as to which task to complete, and two trials of the same task were completed in succession. This enabled contrasts between ERPs elicited by cues on *stay* trials, where the cue on the preceding trial signalled the same retrieval task, and *switch* trials, where the cue differed from that on the preceding trial. A significant difference between the preparatory activities following each cue was evident for switch trials only. These findings diverge from previous outcomes where the activity differentiating cues signalling preparation for episodic or semantic retrieval has been restricted to stay trials. In previous studies the episodic/semantic switching requirement was accompanied by the additional requirement to switch to recover different kinds of contents. This second requirement was minimised here because location information was required for both episodic and semantic judgments. These findings suggest, therefore, that inferences made previously about a consistent neural signature associated with preparation for episodic retrieval do not hold under certain circumstances.

3.2   Experiment 1: Introduction

The notion of retrieval mode was originally proposed by Tulving (1983) as part of the General Abstract Processing System (GAPS) model. Tulving (1983) suggested that in order to engage in episodic memory retrieval one must be in a particular cognitive state. It was suggested that this state allows inputs from the environment to be treated as retrieval cues for past events, and permits 'mental time travel' which enables thinking in subjective space and time (Nyberg et al., 1995; Tulving, 1983; Wheeler et al., 1997).

ERP studies have revealed a pattern of activity that has been regarded as a neural signature of retrieval mode (Düzel et al., 1999; Düzel et al., 2001; Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). This signature comprises relatively more positive-going ERP activity at right-frontal scalp sites when preparing for an episodic memory task, in comparison to a semantic baseline task.

However, when these past studies have manipulated retrieval mode by examining differences between neural activities when people are preparing for either episodic or semantic memory, the content of the memory to be retrieved has also varied between the tasks. For instance, the episodic task in Herron and Wilding (2006a) consisted of retrieving information about the screen position of a word from a prior exposure phase, and the semantic task involved participants recalling whether the object denoted by the word could move by its own accord. As a result, it is unclear whether the ERP activity evident when preparing for episodic retrieval in these studies is indeed an index of retrieval mode, or reflects differences due to the content of what was being retrieved. In order to assess this possibility, the aim of Experiment 1 was to examine indices of preparatory retrieval processing when the contents of the episodic and semantic memory tasks were made similar.

In this experiment, the content was location information for both tasks. For the episodic task participants were required to indicate the screen location of the word from the prior study phase (inside/outside/new). For the 'baseline' semantic task participants were required to indicate the common location of the object depicted (inside/outside/both). Keeping the content matched across the two tasks (both location judgments) allowed for any differences between the ERP activities to be attributed to differential preparatory processing for the episodic task in comparison to the semantic task, rather than to differences between

the contents of retrieval. As in the previous task-switching studies investigating electrophysiological indices of preparation for episodic memory retrieval (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002), here participants were also cued on a trial-by-trial basis, allowing comparisons between ERPs elicited by the cues on switch and stay trials.

If the signature of retrieval mode identified in previous ERP studies is in fact determined by the demands imposed by completion of an episodic versus a non-episodic task, then the preparatory ERP activity should be more positive-going following the episodic task cue in comparison to the semantic task cue at right-frontal scalp sites from approximately 800ms onwards on stay trials. Ruling out the possible confound of the content of retrieval would further substantiate the claim that there is activity that can be referred to as a distinct neural marker for retrieval mode. On the other hand, the absence of significant differences between the critical preparatory ERPs at right-frontal sites would challenge existing functional characterisations.

3.3   Experiment 1: Method

3.3.1   Participants

A sample size of 24 participants was decided *a priori* based on counterbalancing considerations and power analyses for a replication attempt of the effects found previously (see **Appendix A**). Data were collected from 26 participants, and the data from two participants were excluded: one due to excessive EEG artefact and one due to a semantic categorisation score of below 40%. Thus, the data from 24 participants (M_{age} = 21, range = 18-26, 19 female) were included in the analyses. In this and in all subsequent experiments, all participants gave informed consent before participating, and were right-handed, with normal or corrected-to-normal vision. None of the participants had a diagnosis of dyslexia, and they were all native English speakers. At the time of testing none of the participants were using psychotropic medication. In this experiment, participants were paid £10 per hour.

3.3.2   Design

The stimuli were 240 black line drawings of objects, selected from the International Picture Naming Project Database (Szekely et al., 2004). The corresponding name for each object was between three and ten letters in length, the percentage picture naming frequency was above 0.80, and the frequency range was between zero and 7.396 CELEX log

transformed (Szekely et al., 2004). The objects were presented on a monitor with a white background, positioned one metre directly in front of participants. The stimuli subtended maximum visual angles of 5.4˚ vertically and 8.5˚ horizontally at study. At test, objects were presented in the centre of the screen subtending maximum visual angles of 1.6˚ vertically and 1.7˚ horizontally.

The objects were classified into one of three semantic categories, according to where they were commonly found: inside, outside or both. An object was classified as 'inside' if it was usually found inside, and it was classified as 'outside' if it was usually found outside. The object was classified as 'both' if it could commonly be found both inside and outside. There were 80 objects in each semantic category, and for this classification the mean inter-rater reliability of three raters was 0.72. The experiment comprised five study-test cycles, and the 80 stimuli from each semantic category were randomly assigned to one of five lists. Thus, each list contained 48 objects with 16 from each semantic category (inside/outside/both). Two additional practice blocks, half the length of the other five study-test blocks, were formed and used to familiarise participants with the experiment demands.

At study, 24 of the objects were either presented inside or outside an abstract outline of a building (**Figure 6**). They were displayed in one of eight randomised locations (four inside, four outside). Half of the objects were presented inside and half of the objects were presented outside, and this was counterbalanced across individuals (for all experiments in this thesis). Participants were asked to indicate whether the object appeared inside or outside, and to make a response via button press with their middle or index fingers respectively. The hand used was counterbalanced across participants.
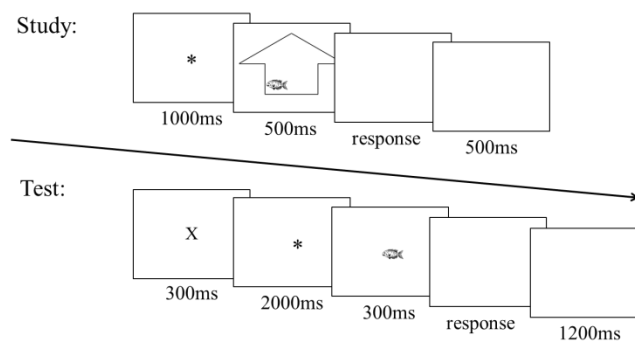
Figure 6. Schematic illustration of a trial sequence and timing at study and test.

At test, the 24 objects from the preceding study phase were randomly intermixed with 24 unstudied objects. Each test object was preceded by one of two preparatory cues presented in the centre of the screen. These cues indicated which task participants were to subsequently complete. A capital 'X' directed participants to prepare for the episodic task, where they were required to retrieve the prior study location of the object. A capital 'O' directed the participants to prepare for the semantic task. This task required identification of the common location of the object depicted, regardless of the study phase. Each cue-type was always presented for two consecutive trials. Trials where the cue signalled a different task to the previous trial are referred to as *switch* trials, whereas trials where the cue signalled the same task as the preceding trial are referred to as *stay* trials. Whether the object appeared on a switch or stay trial, its task status (episodic/semantic), and old/new status, were counterbalanced across participants. During the test phase, responses were made using the same fingers as at study, with the addition of the index finger of the other hand to indicate 'new' or 'both', for the episodic and the semantic task respectively.

3.3.3   Procedure

At study, a fixation asterisk was presented for 1000 milliseconds (ms), followed by an object (presented inside or outside the building outline) for 500ms. The monitor was then blank until a response was made, and remained blank for 500ms following each response before the next trial began. At test, the preparatory cue ('X' or 'O') appeared for 300ms, followed by a fixation asterisk for 2000ms. Following a 'X' cue, participants were required to prepare to retrieve information about whether the object appeared inside or outside at study, or whether the object was new (response: inside/outside/new). Following the 'O' cue, participants were required to prepare to identify where the object was most commonly found: inside or outside, or both inside and outside (response: inside/outside/both). An object was then presented in the centre of the screen for 300ms. The monitor was then blank until a response was made, and remained blank for a further 1200ms before the next trial began (**Figure 6**). Participants were asked to respond as quickly and as accurately as possible. Trials on which responses were faster than 300ms or slower than 4000ms were counted as errors and excluded from the behavioural analyses (0.9% of the trials).

3.3.4 EEG acquisition

General EEG acquisition details are contained in the **General methods**. ERPs were recorded time-locked to the preparatory cues as well as to the items following the cues. For the preparatory cue-elicited ERPs the epoch length was 2500ms including a 200ms pre-stimulus baseline. Item-elicited ERPs were obtained by extracting EEG epochs of 1400ms, including a 200ms baseline (see ***Appendix B***).

3.4 Experiment 1: Results

3.4.1 Frequentist statistics

In this and in all subsequent experiments, the first trial in each test block was removed from analyses, as it is neither a switch nor a stay trial. On average, 84% of the trials contributed to the ERP cue data in this experiment for each participant. For the cue data the mean trial numbers contributing to the ERPs (ranges in parenthesis) were: episodic switch = 50 (34-58), episodic stay = 52 (32-60), semantic switch = 49 (31-58), semantic stay = 51 (31-60). ERPs elicited by the test items following the preparatory cues are included in ***Appendix B***.

3.4.1.1 Behavioural analyses

During the study phases participants correctly responded 'inside' or 'outside' on 96% of trials. **Table 4** shows the response accuracy data for the test phases.

Table 4. Response accuracies for each task on switch and stay trials in Experiment 1. Standard deviations are in parentheses.

	Switch	Stay
Episodic task:		
Old/new discrimination (*Pr*)	0.58 (0.19)	0.65 (0.17)
P(correct source)	0.75 (0.13)	0.81 (0.11)
Correct rejection	0.73 (0.19)	0.78 (0.17)
Semantic task:		
Correct classification	0.73 (0.08)	0.73 (0.07)

Discrimination scores (discrimination index: $Pr = p$(hit) - p(false alarm), Snodgrass & Corwin, 1988) for the episodic task were above zero for both trial-types (switch: 0.58, stay: 0.65) and higher on stay trials than on switch trials ($t(23) = 2.56$, $p < 0.05$, $d_z = 0.52$, 70% CL).

The conditional probabilities (given a correct old response) of correct source judgments collapsed across the inside/outside dimension were reliably above chance in both cases (switch: $t(23) = 9.16$, $p < 0.001$, $d_z = 1.87$, 97% CL; stay: $t(23) = 13.33$, $p < 0.001$, $d_z = 2.72$, 99.7% CL). Performance was superior on stay trials ($t(23) = 2.77$, $p < 0.05$, $d_z = 0.57$, 71% CL). For the semantic task, the probability of classifying the item according to the modal rating given by the original raters was equivalent for switch and stay trials.

A 2x2x2 ANOVA was conducted on the reaction times (see **Table 5**) for response accuracy categories (correct episodic source or semantic categorisation: when old/when new) separated according to task (episodic/semantic) and trial-type (switch/stay). Main effects of trial-type ($F(1, 23) = 16.32$, $p = 0.001$, $d_z = 0.82$, 80% CL), and response accuracy ($F(1, 23) = 18.53$, $p < 0.001$, $d_z = 0.88$, 81% CL), were moderated by a trial-type by response accuracy by task interaction ($F(1, 23) = 4.40$, $p < 0.05$, $\eta_p^2 = 0.16$) (there was also an interaction between task and response accuracy: $F(1, 23) = 5.36$, $p < 0.05$, $\eta_p^2 = 0.19$, and a trend between trial-type and response accuracy: $F(1, 23) = 3.97$, $p = 0.058$, $\eta_p^2 = 0.15$).

Looking at each task separately there were main effects of trial-type ($F(1, 23) = 6.81$, $p < 0.05$, $d_z = 0.53$, 70% CL) and response accuracy ($F(1, 23) = 11.70$, $p < 0.05$, $d_z = 0.70$, 76% CL) for the episodic task: reflecting slower responses for switch in comparison to stay, and for correct when old than correct when new responses. For the semantic task there was also a main effect of trial-type ($F(1, 23) = 12.00$, $p < 0.05$, $d_z = 0.71$, 76% CL) and a trend for response accuracy ($F(1, 23) = 2.97$, $p = 0.098$, $d_z = 0.35$, 64% CL), however, these were moderated by an interaction between these factors ($F(1, 23) = 6.30$, $p < 0.05$, $\eta_p^2 = 0.22$). This reflects significantly slower responses for switch than stay trials for correct when old ($p < 0.001$) but not correct when new responses, and for correct when old than correct when new responses for switch trials ($p < 0.05$) but not for stay trials.

Table 5. Reaction times (ms) for each task on switch and stay trials in Experiment 1. Standard deviations are in parentheses.

	Switch	Stay
Episodic task:		
Correct source	1509 (478)	1412 (381)
Correct new	1327 (356)	1226 (287)
Semantic task:		
Correct old	1483 (368)	1342 (338)
Correct new	1385 (313)	1367 (397)

3.4.1.2 ERP analyses: cue data

The ERPs elicited by the two cues indicating which task to complete were analysed over an 800 to 1900ms time window. Previous researchers have shown differences between the preparatory activity associated with episodic and semantic retrieval within this epoch (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). In keeping with the analysis approach adopted in previous ERP studies in which electrophysiological indices of retrieval mode were investigated (Herron & Wilding, 2004, 2006a), the initial analysis included 12 sites distributed over fronto-central regions (F3/F4, F5/F6, F7/F8, C3/C4, C5/C6, T7/T8). Consequently, the initial ANOVA was conducted incorporating the factors of: cue-type (episodic/semantic), trial-type (switch/stay), location in the anterior-posterior plane (anterior/central), hemisphere (left/right), and site (inferior/mid-lateral/superior). In this and in all comparable analyses in this thesis, only outcomes involving the factor of cue-type are reported.

Figure 7 shows the grand averaged ERP waveforms for each cue-type at right anterior and central sites, separated for switch and stay trials. Scalp maps depicting the differences between the scalp distributions of the ERPs associated with the cue-types are shown in **Figure 8**, again separated according to switch and stay trials (A & B). These figures demonstrate a greater relative positivity at right-frontal sites for the semantic task on switch trials. Moreover, there are some divergences between the ERPs elicited by the two cue-types at central electrode locations on switch and on stay trials.

Switch
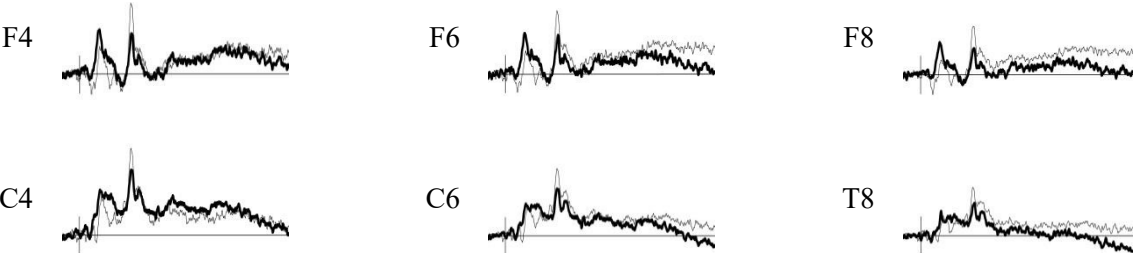
Stay
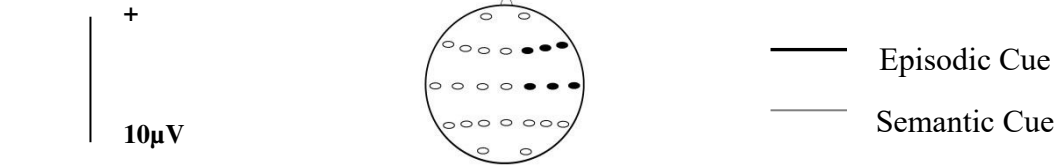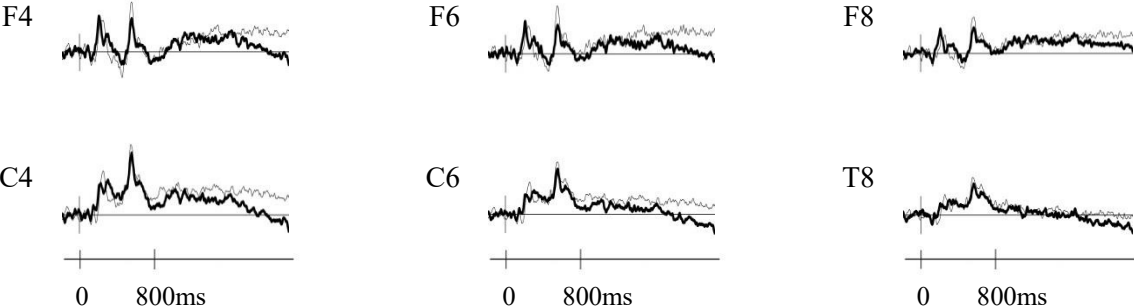
Figure 7. Grand averaged ERPs for Experiment 1 separated according to trial-type (switch/stay) and cue-type (episodic/semantic) for right anterior (F4, F6, F8) and right central electrode sites (C4, C6, T8).

A. Episodic Switch – Semantic Switch   **B.** Episodic Stay – Semantic Stay

800 ms - 1900 ms   800 ms - 1900 ms

-1.68 µV   0 µV   1.38 µV   -1.59 µV   0.44 µV
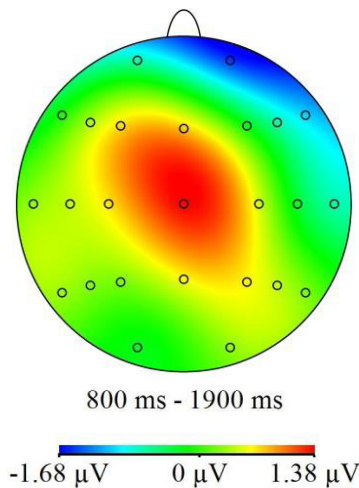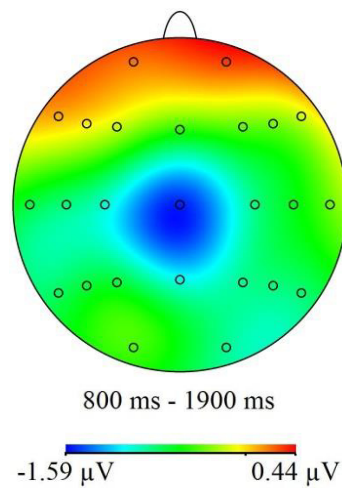
Figure 8. Topographic maps for Experiment 1: the differences between the scalp distributions of the ERPs associated with the episodic and semantic cues on switch (A) and stay (B) trials from 800 to 1900ms. The scale below each map denotes the voltage range (µV) of the differences between conditions.

The initial analysis revealed an interaction involving the factors of cue-type, trial-type, anterior-central dimension, hemisphere, and site ($F(1.7, 39.3) = 4.35$, $p < 0.05$, $\eta_p^2 = 0.16$). There were also lower order interactions involving: cue-type and hemisphere ($F(1, 23) = 4.62$, $p < 0.05$, $\eta_p^2 = 0.17$), cue-type, trial-type, and the anterior-central dimension ($F(1, 23) = 4.37$, $p < 0.05$, $\eta_p^2 = 0.16$) and cue-type, the anterior-central dimension, and hemisphere ($F(1, 23) = 6.70$, $p < 0.05$, $\eta_p^2 = 0.23$).

In light of the interactions involving cue- and trial-type, separate ANOVAs were carried out for switch and stay trials, and in both cases reliable differences according to cue-type were evident. For switch trials the analyses revealed interactions between: cue-type, anterior-central, hemisphere and site ($F(1.7, 39.4) = 4.66$, $p < 0.05$, $\eta_p^2 = 0.17$), as well as cue-type and hemisphere ($F(1, 23) = 8.41$, $p < 0.05$, $\eta_p^2 = 0.27$), and a trend between cue-type and site ($F(1.2, 26.6) = 3.87$, $p = 0.055$, $\eta_p^2 = 0.15$). Follow up ANOVAs were subsequently carried out separately for the anterior and central sites and revealed reliable differences at anterior sites only, comprising an interaction between cue-type and hemisphere ($F(1, 23) = 8.31$, $p < 0.05$, $\eta_p^2 = 0.27$) and an interaction between cue-type and site ($F(1.3, 29.6) = 4.00$, $p < 0.05$, $\eta_p^2 = 0.15$). The cue-type by hemisphere interaction reflects more positive-going activity for the semantic task in comparison to the episodic task over right hemisphere sites,

and more positive-going activity for the episodic task than the semantic task over left hemisphere sites. The interaction involving site reflects the greater relative positivity for the semantic task at inferior and mid-lateral sites, but more positive-going ERP activity for the episodic task at superior sites. At central sites there were only trends for interactions between cue-type, hemisphere and site ($F(1.3, 29.9) = 3.64$, $p = 0.056$, $\eta_p^2 = 0.14$), cue-type and hemisphere ($F(1, 23) = 4.21$, $p = 0.052$, $\eta_p^2 = 0.15$), and cue-type and site ($F(1.2, 28.7) = 2.79$, $p = 0.099$, $\eta_p^2 = 0.10$).

On stay trials an interaction was evident between cue-type and the anterior-central dimension ($F(1, 23) = 4.31$, $p < 0.05$, $\eta_p^2 = 0.16$), as well as a trend between cue-type, anterior central dimension, and hemisphere ($F(1, 23) = 3.61$, $p = 0.07$, $\eta_p^2 = 0.14$). Separate follow ups at anterior and central sites revealed no reliable outcomes. The higher order cue-type by anterior-central interaction reflects the greater relative positivity for the semantic task across central electrode sites, with more comparable measures for each task cue-type over anterior electrode sites.

3.4.2 Bayesian statistics

3.4.2.1 ERP analyses: replication of right-frontal positivity during preparation for episodic memory retrieval

Bayes Factors (BFs) were calculated in order to investigate whether the ERP results provide support for the null hypothesis, according to Bayesian statistics (Dienes, 2011; Verhagen & Wagenmakers, 2014) (see **General methods**). A BF compares two hypotheses, and for this experiment they were: the alternative hypothesis that there would be more positive-going ERP activity at right-frontal scalp sites when preparing for the episodic task in comparison to the semantic task, and the null hypothesis that this would not be the case. Computation of replication success requires the number of participants and the t value for the main effect of interest, in an original and replication experiment (**Table 6**).

t values for the replication experiments were calculated using the same analysis strategies and sites as in the previous studies reporting main effects of cue-type (Herron & Wilding, 2004, 2006a). A main effect of cue-type was not reported in the Morcom and Rugg (2002) study. Instead, interactions involving the factor of cue-type were reported for stay trials (cue-type by site, cue-type by hemisphere). In addition, in the Morcom and Rugg

(2002) study a different electrode montage was used. Thus, mean amplitudes for the same electrode sites could not be computed, and BFs were not calculated. For Herron and Wilding (2004), t values were calculated using the six right-frontal and central sites for stay trials, for the epoch 800-1900ms. There were two episodic tasks: operation and location, and one baseline semantic task. For Herron and Wilding (2006a) the t value was calculated using the three right-frontal sites for stay trials, and the mean amplitude measures from 800ms until the retrieval cues were presented (4000ms). There was one episodic location task, and one baseline semantic task.

Table 6. t values and sample sizes (N) from the original studies demonstrating a greater right-frontal positivity during preparation for episodic than for semantic memory retrieval, and for the replication attempt (Experiment 1).

Study	t value	N
Herron and Wilding (2004):		20
Main effect of cue-type (operation/semantic)	2.09	
Main effect of cue-type (location/semantic)	2.86	
Herron and Wilding (2006a):		16
Main effect of cue-type (location/semantic)	2.47	
Experiment 1:		24
Analysis strategy as 2004 study	-0.79	
Analysis strategy as 2006a study	-1.00	

As a replication of the episodic retrieval mode operation task effect in Herron and Wilding (2004) the $BF_{r0} = 0.12$ (**Figure 9A**). For a replication of the location task effect the $BF_{r0} = 0.04$ (**Figure 9B**). These BFs indicate that the data provide substantial and strong evidence, respectively, for the null relative to the alternative hypothesis. For a replication of the location task effect in Herron and Wilding (2006a) the $BF_{r0} = 0.05$ (**Figure 9C**). Again, this indicates that the data provide strong evidence in favour of the null hypothesis. Thus, the outcomes suggest strongly that the right-frontal preparatory ERP index reported in a number of previous studies is not present here ($BF_{ave} = 0.07$, strong evidence for the null hypothesis).

Herron and Wilding (2004)   **Herron and Wilding (2006a)**

A. Operation   **B.** Location   **C.** Location
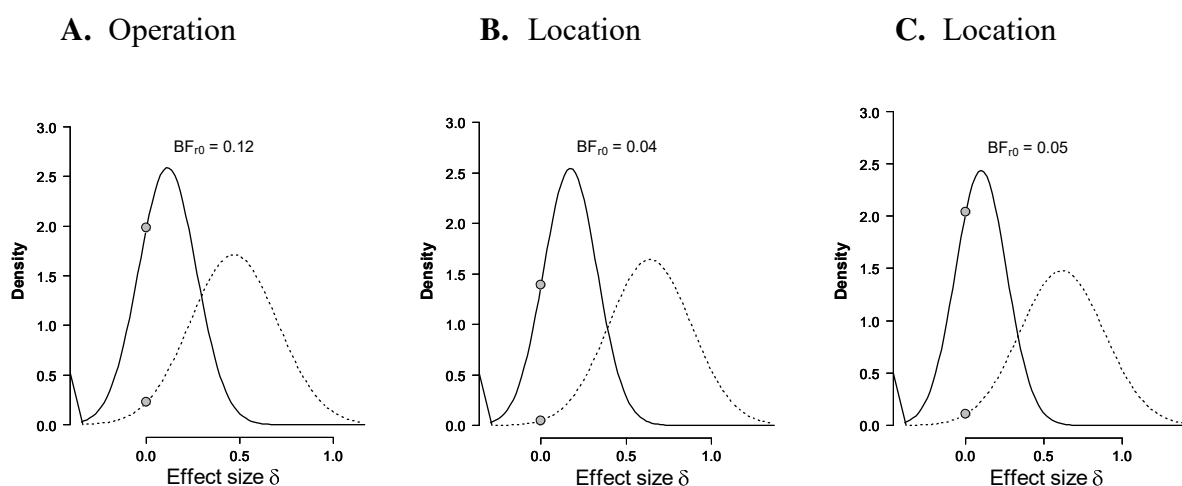
Figure 9. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the right-frontal positivity identified previously during preparation for episodic memory retrieval. In each panel, the dotted lines represent the posterior from the original study (Herron & Wilding, 2004, A & B; Herron & Wilding, 2006a, C), which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Experiment 1) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

In addition, a Bayesian replication test (Verhagen & Wagenmakers, 2014) was also carried out to explore whether the ERP results of Herron and Wilding (2006a) provide support for the alternative hypothesis as a replication of the effects found in Herron and Wilding (2004). For Herron and Wilding (2006a) as a replication of the episodic retrieval mode operation task effect in Herron and Wilding (2004) the $BF_{r0} = 9.66$ (**Figure 10A**). For a replication of the location effect the $BF_{r0} = 10.47$ (**Figure 10B**). Thus, the BFs provide substantial and strong evidence in favour of the alternative relative to the null hypothesis. The data are approximately ten times more likely under the alternative than the null hypothesis, demonstrating the robustness of the effects previously reported in the literature.

A. Herron & Wilding 2006a: for 2004 'operation' effect
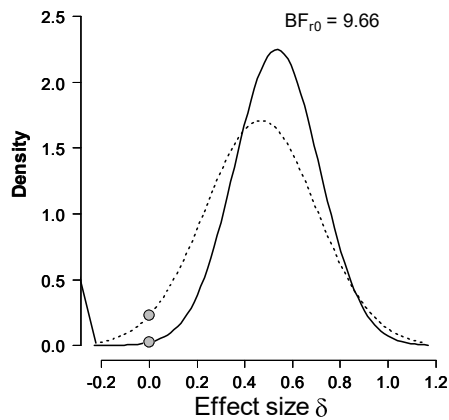
B. Herron & Wilding 2006a: for 2004 'location' effect
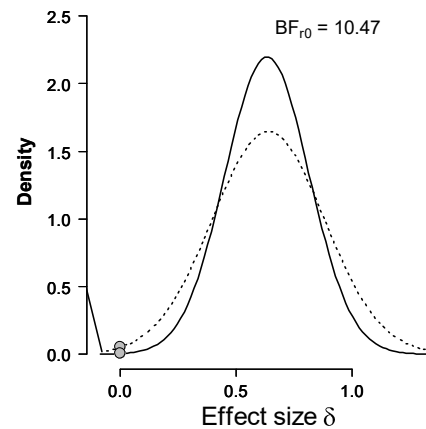
Figure 10. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the right-frontal positivity identified previously during preparation for episodic memory retrieval. In each panel, the dotted lines represent the posterior from the original study (Herron & Wilding, 2004), which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Herron & Wilding, 2006a) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

3.5 Experiment 1: Discussion

The aim of this experiment was to investigate the ERP activities when participants prepared to complete an episodic memory task in comparison to a semantic memory task, whilst keeping the content of the tasks highly similar. In previous studies an ERP modulation was identified that was proposed to index retrieval mode (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). This index was obtained in tasks where people were cued trial-by-trial to prepare for either episodic or semantic retrieval. In addition, it was observed for stay trials only, where the preparatory cue indicated preparation for episodic retrieval for the second trial in a row. However, in these previous studies the contents to be retrieved varied between the episodic and baseline semantic tasks, leaving open the possibility that this ERP signature reflects to an unknown extent this difference in content, rather than only the requirement to recover episodic information.

In this experiment, accuracy switch costs (old/new discrimination and source accuracy for the episodic task), as well as reaction time costs, were observed. In addition, the preparatory ERP activity for the tasks diverged. However, this was reliable only on switch trials and consisted of more positive-going activity for the semantic task over right hemisphere anterior scalp sites, and more positive-going activity for the episodic task over left hemisphere anterior scalp sites.

Reaction time switch costs have been evident in previous task-switching studies investigating preparation for episodic memory retrieval (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002), and were replicated in this experiment: participants were faster on stay trials in comparison to switch trials. These reaction time switch costs, in the absence of costs for the accuracy of the memory judgments led Herron and Wilding (2004) to propose that adopting a task-relevant set facilitates the time course of retrieval processing, possibly "maximising the efficiency of search operations, or the efficiency with which information is processed" (Herron & Wilding, 2004, p. 1560). However, in this experiment old/new discrimination was superior on stay trials, whereas classification accuracy in the semantic task showed no comparable switch cost. Source accuracy was also superior on stay trials in comparison to switch trials.

These outcomes raise the question of why similar accuracy costs have not been reported in previous studies. The primary difference between the experiment design employed here and those employed previously is the similarity between the contents that have to be retrieved in the episodic and semantic tasks. One possibility is that the similarity between contents resulted in a high degree of interference between information being accessed, and this interference impacted negatively on performance measures in the episodic task.

The ERP divergence in this experiment was also not consistent with what has been observed across a number of previous studies (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). The ERPs elicited by the episodic and semantic preparatory cues diverged on the switch trials, rather than the stay trials, with more positive-going activity for the semantic than the episodic task over the right-frontal scalp sites. Additional Bayesian analyses gave BFs that provided strong support for the null relative to the alternative hypothesis: there was no evidence for the putative retrieval mode ERP index identified previously on stay trials.

77

These data, therefore, raise questions about the generality of the claim that there is a specific ERP index of retrieval mode with predictable temporal and spatial properties.

In summary, the right-frontal ERP modulation reported on stay trials in previous studies of preparatory retrieval processing was not observed in this experiment. This is the first study in which attempts were made to equate the kinds of content upon which episodic and semantic memory judgments might be made. The findings suggest that there is not a consistent neural index of retrieval mode, and emphasise that memory contents as well as memory systems need to be considered when characterising how people prepare to recover information from long-term memory.

4. CHAPTER 4. Experiment 2: content unmatched

4.1 Experiment 2: Abstract

One possibility is that the findings in Experiment 1, and the ways in which they diverged from those in previous studies, were due to attempts at matching the content of the episodic and semantic memory tasks. As in Experiment 1, in this experiment comparisons were made between the ERPs elicited by two preparatory cues signalling that participants should prepare for different retrieval tasks. One cue signalled preparation for episodic retrieval: identify the location in which an object had been seen in a prior study phase. However, here, the other cue signalled that participants were to prepare to identify the size of the object depicted (semantic retrieval task). Participants were cued trial-by-trial as to which task to complete, and two trials of the same task were completed in succession. There were no significant differences between the cue-types for either trial-type (switch/stay). These findings diverge from previous outcomes where activity differentiating cues signalling preparation for episodic or semantic retrieval has been evident on stay trials. Additionally, the right-frontal semantic task positivity obtained on switch trials in Experiment 1 was not replicated in this experiment, and neither was the source accuracy switching cost. These findings add further support to the claim that inferences made previously about consistent neural signatures associated with preparation for episodic retrieval do not hold under certain circumstances.

4.2   Experiment 2: Introduction

In task-switching studies in which electrophysiological indices of preparation for episodic memory retrieval have been investigated, an index was identified that was proposed to capture the theoretical notion of retrieval mode (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). ERP activity was more positive-going on stay trials following cues indicating preparation for episodic tasks in comparison to cues indicating preparation for baseline semantic tasks. This activity was evident at right-frontal scalp sites from approximately 800 to 1900ms.

In these previous experiments, however, the content of retrieval differed across the tasks. Thus, it is possible that the divergences in ERP activities observed previously could reflect these content-specific differences. Experiment 1 of this thesis was designed in order to rule out this potential confound. Electrophysiological indices of preparation for episodic memory retrieval were investigated whilst attempting to equate the content of the episodic and baseline semantic task judgments. Participants were required to retrieve 'location' information for both tasks, either: (i) the study location (episodic task: inside/outside/new) or (ii) the common location of the object in everyday life (baseline semantic task: inside/outside/both).

In Experiment 1, activity differentiating the cue-types was restricted to switch trials, and was more negative-going following the episodic task cue in comparison to the semantic task cue over right hemisphere anterior sites. Thus, with content matched across the tasks there was no evidence for the putative retrieval mode ERP index (where activity differentiating the cues was restricted to stay trials, and more positive-going for the episodic task). The additional requirement to switch to recover different kinds of content, evident in the previous studies, was minimised in Experiment 1. Therefore, the results of Experiment 1 suggest that the effects found previously were indices of retrieving certain kinds of content-specific information, rather than reflecting the general requirement to recover episodic information.

Experiment 2 was designed to test the proposal that the absence of an identifiable retrieval mode ERP index in Experiment 1 (and the presence of other effects) was due to matching the contents for retrieval. In Experiment 2 the episodic task remained the same as in

Experiment 1 (prior study location judgment), but the semantic task required the retrieval of a different content (a relative size judgment). For this task participants were required to judge the size of the object depicted. Thus, as in the previous experiments in the literature, for Experiment 2 the content of the memory decision was not matched across the tasks. It was anticipated that the effects obtained in the previous content unmatched studies (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002) would be replicated here, comprising a greater relative positivity for the episodic than for the semantic task at right-frontal locations on stay trials.

In addition, in Experiment 1 there were accuracy switch costs which have not often been observed in task-switching memory experiments (Herron & Wilding, 2004; Morcom & Rugg, 2002). It is possible that matching the contents across the tasks in Experiment 1 is the reason for the additional switch costs seen there. Thus, as in previous task-switching memory studies with unmatched content, the absence of switching costs for the accuracy of episodic judgments was anticipated.

4.3   Experiment 2: Method

4.3.1   Participants

Data were collected and analysed from 32 participants (M_{age} = 22, range = 18-28, 21 female) based on the Bayesian Stopping Rule (Dienes, 2011) (see **General methods &** ***Appendix A***). Participants were paid £10 per hour.

4.3.2   Design

The stimuli were the same 240 black line drawings of objects (Szekely et al., 2004) used in Experiment 1. The objects were presented following the same protocol as in Experiment 1. However, in this experiment the objects were classified into one of three semantic categories according to the size of the object depicted: smaller than a lunchbox, larger than a lunchbox but smaller than a suitcase, larger than a suitcase. There were approximately a third of the objects in each semantic category (88, 62, and 90, respectively). For the semantic classification, the mean inter-rater reliability of three raters was 0.66. The experiment comprised five study-test cycles, and for each cycle each of the stimuli from each semantic category were randomly assigned to one of five lists. Thus, each list contained 48 objects: 17/18 small, 12/13 medium, and 18 large objects. Two additional practice blocks,

half the length of the other five study-test blocks, were formed and used to familiarise participants with the experiment demands.

At study, 24 of the objects were either presented inside or outside an abstract outline of a building, and participants were required to indicate whether the object appeared inside or outside as in Experiment 1 (see **Figure 6**. Experiment 1. p. 66).

At test, the 24 objects from the preceding study phase were randomly intermixed with 24 unstudied objects. As in Experiment 1, each test object was preceded by one of two preparatory cues (X/O) presented in the centre of the screen, indicating which task participants were to subsequently complete. However, in this experiment the capital 'O' cue signalled that the semantic task that followed required the retrieval of information about the size of the object depicted. As in Experiment 1, each cue-type was always presented for two consecutive trials only. Item status was counterbalanced across participants in the same way as in Experiment 1. During the test phase responses were made using the same fingers as at study, with the addition of the index finger of the other hand to indicate 'new' or 'larger than both', for the episodic and the semantic task respectively. The hands used for the judgments were counterbalanced across participants.

4.3.3 Procedure

The procedure followed the same trial sequence as in Experiment 1 (see **Figure 6**. Experiment 1. p. 66). However, following the 'O' cue for the semantic task, participants were required to prepare to retrieve information about whether the object depicted would fit inside a lunchbox, a suitcase, or whether it was larger than both of these containers (<lunchbox, >lunchbox<suitcase, >both). Trials on which responses were faster than 300ms or slower than 4000ms were counted as errors and excluded from the behavioural analyses (0.6% of the trials).

4.3.4 EEG acquisition

The EEG was recorded following the same protocol as in Experiment 1 (see **General methods**).

4.4   Experiment 2: Results

4.4.1   Frequentist statistics

Unless otherwise stated, the same analysis procedures were followed as for Experiment 1. On average, 86% of the switch and stay trials contributed to the ERP cue data in this experiment for each participant. The mean trial numbers contributing to the ERPs (ranges in parentheses) were: episodic switch = 50 (27-57), episodic stay = 53 (31-60), semantic switch = 51 (34-58), semantic stay = 52 (29-59). ERPs elicited by the test items following the preparatory cues are included in *Appendix B*.

4.4.1.1   Behavioural analyses

During the study phases participants correctly responded 'inside' or 'outside' on 98% of trials. **Table 7** shows the response accuracy data for the test phases.

Table 7. Response accuracies for each task on switch and stay trials in Experiment 2. Standard deviations are in parentheses.

	Switch	Stay
Episodic task:		
Old/new discrimination (*Pr*)	0.61 (0.20)	0.68 (0.23)
P(correct source)	0.76 (0.13)	0.77 (0.11)
Correct rejection	0.75 (0.19)	0.83 (0.17)
Semantic task:		
Correct classification	0.75 (0.08)	0.75 (0.08)

Discrimination scores (discrimination index: $Pr = p$(hit) - p(false alarm), Snodgrass & Corwin, 1988) for the episodic task were above zero for both trial-types (switch: 0.61, stay: 0.68) and higher on stay trials than on switch trials ($t(31) = 2.95$, $p < 0.05$, $d_z = 0.52$, 70% CL).

The conditional probabilities of correct source judgments collapsed across the inside/outside dimension were reliably above chance in both cases (switch: $t(31) = 11.70$, $p < 0.001$, $d_z = 2.07$, 98% CL; stay: $t(31) = 13.31$, $p < 0.001$, $d_z = 2.35$, 99% CL). For the semantic task, the probability of classifying the item according to the modal rating given by

the original raters was equivalent for switch and stay trials. Source accuracies in the episodic task and the semantic task accuracies were not statistically different across trial-types.

A 2x2x2 ANOVA was conducted on the reaction times (see **Table 8**) for response accuracy categories (correct episodic source or semantic categorisation: when old/when new) separated according to task (episodic/semantic) and trial-type (switch/stay). Main effects of trial-type ($F(1, 31) = 11.97$, $p < 0.05$, $d_z = 0.71$, 76% CL), task ($F(1, 31) = 10.32$, $p < 0.05$, $d_z = 0.66$, 74% CL) and response accuracy ($F(1, 31) = 44.44$, $p < 0.001$, $d_z = 1.36$, 91% CL), were moderated by an interaction between all three factors ($F(1, 31) = 5.26$, $p < 0.05$, $\eta_p^2 = 0.15$) (there was also a task by response accuracy interaction: $F(1, 31) = 10.48$, $p < 0.05$, $\eta_p^2 = 0.25$).

Looking at each task separately there were main effects of trial-type ($F(1, 31) = 9.74$, $p < 0.05$, $d_z = 0.64$, 74% CL) and response accuracy ($F(1, 31) = 26.58$, $p < 0.001$, $d_z = 1.05$, 85% CL) for the episodic task, moderated by a trial-type by response accuracy interaction ($F(1, 31) = 5.87$, $p < 0.05$, $\eta_p^2 = 0.16$). This interaction reflects slower responses for switch than stay trials for correct when new ($p < 0.001$) but not correct when old responses. For the semantic task there was a main effect of response accuracy ($F(1, 31) = 18.74$, $p < 0.001$, $d_z = 0.88$, 81% CL) reflecting significantly slower correct when old than correct when new responses.

Table 8. Reaction times (ms) for each task on switch and stay trials in Experiment 2. Standard deviations are in parentheses.

	Switch	Stay
Episodic task:		
Correct source	1540 (364)	1539 (369)
Correct new	1351 (250)	1227 (277)
Semantic task:		
Correct old	1375 (171)	1339 (238)
Correct new	1293 (176)	1271 (224)

As in Experiment 1, the ERPs elicited by the two cues indicating which task to complete were analysed over an 800 to 1900ms time window, and the initial analysis included the same 12 sites distributed over fronto-central regions (F3/F4, F5/F6, F7/F8, C3/C4, C5/C6, T7/T8).

Figure 11 shows the grand-averaged ERP waveforms for each cue-type at right anterior and central sites, separated for switch and stay trials. Scalp maps depicting the differences between the scalp distributions of the ERPs associated with the cue-types are shown in **Figure 12**, again separated according to switch and stay trials (A & B).
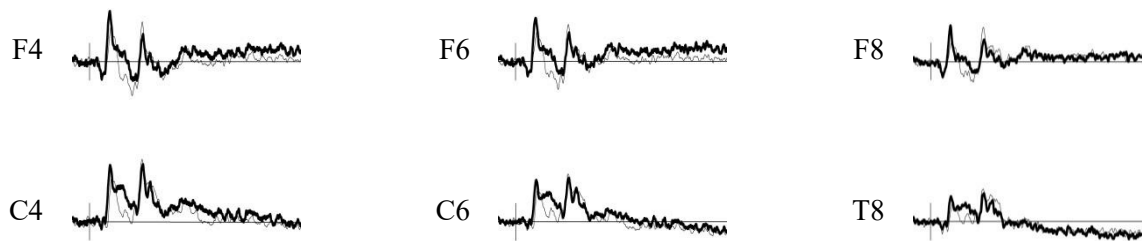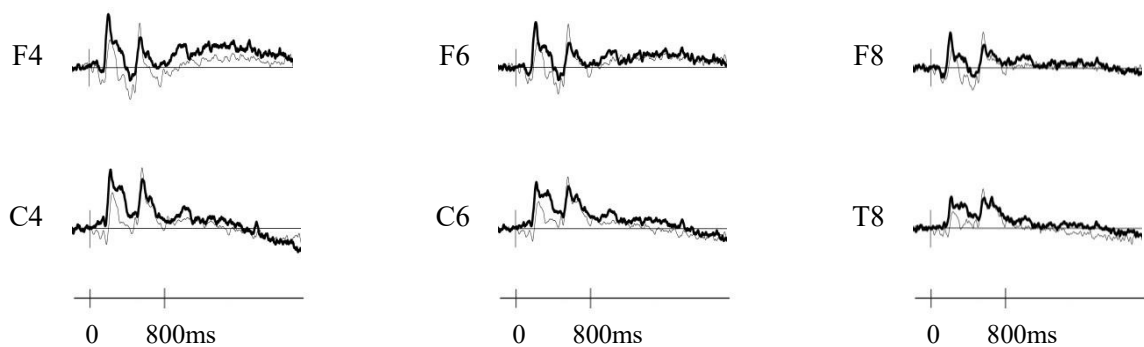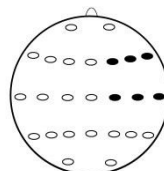
Switch

Stay

Figure 11. Grand averaged ERPs for Experiment 2 separated according to trial-type (switch/stay) and cue-type (episodic/semantic) for right anterior (F4, F6, F8) and right central electrode sites (C4, C6, T8).

A. Episodic Switch – Semantic Switch   **B.** Episodic Stay – Semantic Stay

800 ms - 1900 ms   800 ms - 1900 ms

-0.19 μV   0.77 μV   0.19 μV   0.96 μV
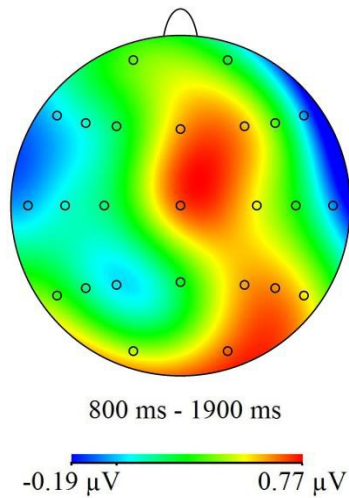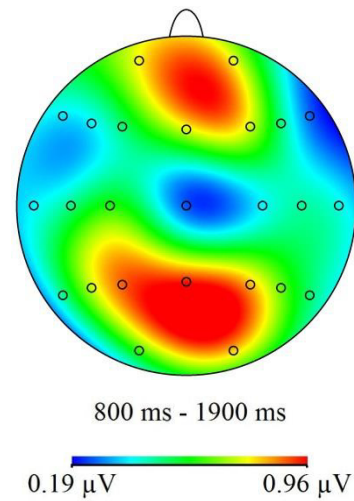
Figure 12. Topographic maps for Experiment 2: the differences between the scalp distributions of the ERPs associated with the episodic and semantic cues on switch (A) and stay (B) trials from 800 to 1900ms. The scale below each map denotes the voltage range (μV) of the differences between conditions.

The initial ANOVA was conducted incorporating the factors of cue-type (episodic/semantic), trial-type (switch/stay), location in the anterior-posterior plane (anterior/central), hemisphere (left/right), and site (inferior/mid-lateral/superior). This revealed trends only: towards a main effect of cue-type ($F(1, 31) = 3.03$, $p = 0.092$, $d_z = 0.31$, 62% CL) and an interaction between cue-type and site ($F(1.4, 44.7) = 3.05$, $p = 0.073$, $\eta_p^2 = 0.09$). Separate exploratory ANOVAs for switch and stay trials, and anterior and central sites, prompted by the findings in Experiment 1, revealed no reliable effects involving cue-type.

4.4.2   Bayesian statistics

4.4.2.1   Behavioural analyses: replication of accuracy switch costs

A Bayesian replication test (Dienes, 2011; Verhagen & Wagenmakers, 2014) was carried out to investigate whether the behavioural results from Experiment 2 provide evidence for a replication of the accuracy switch costs in Experiment 1. The number of participants and the *t* values for the main effects of interest, in the original experiment (1) and replication experiment (2) are shown in **Table 9**.

Table 9. *t* values and sample sizes (*N*) for the accuracy switch costs in Experiment 1, and for the replication attempt (Experiment 2).

Study	*t* value	*N*
Experiment 1:		24
Old/new discrimination switch cost	-2.56	
Source accuracy switch cost	-2.77	
Experiment 2:		32
Old/new discrimination switch cost	-2.95	
Source accuracy switch cost	-0.38	

For Experiment 2 as a replication of the old/new discrimination switch cost in Experiment 1 the $BF_{r0} = 31.88$ (**Figure 13A**). This BF indicates that the data provide very strong evidence for the alternative hypothesis, in line with the outcomes for Experiment 1. For a replication of the source accuracy switch cost the $BF_{r0} = 0.15$ (**Figure 13B**). This BF indicates that the data provide substantial evidence for the null relative to the alternative hypothesis, representing a failure to replicate the Experiment 1 findings.

A. Old/new discrimination　　　　　　　　**B.** Source
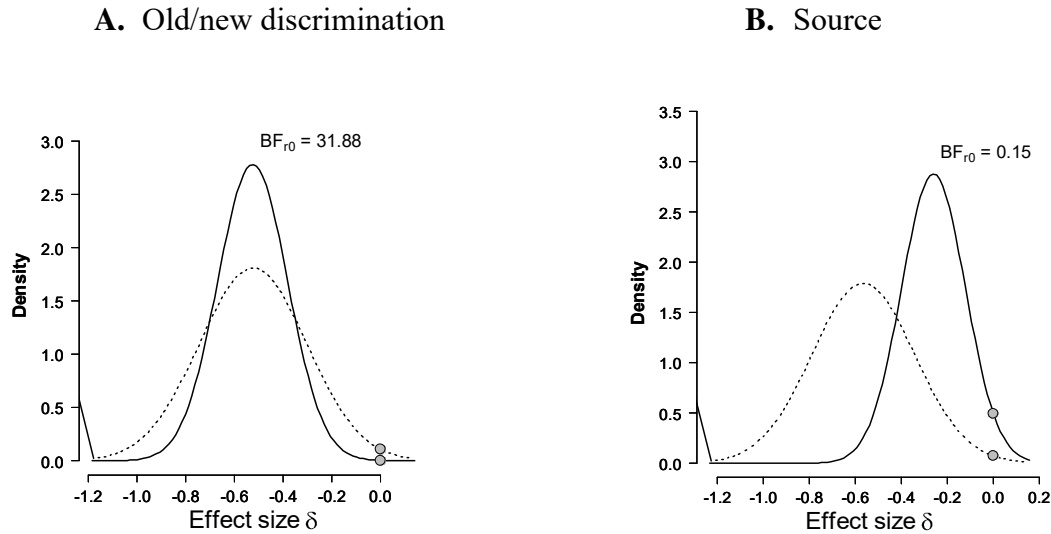
Figure 13. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the accuracy switch costs (A. Old/new discrimination; B. Source) in Experiment 1. In each panel, the dotted lines represent the posterior from the original study, which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Experiment 2) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

4.4.2.2 ERP analyses: replication of right-frontal positivity during preparation for episodic memory retrieval

A Bayesian replication test (Dienes, 2011; Verhagen & Wagenmakers, 2014) was carried out in order to investigate whether the ERP results from Experiment 2 provide support for the null or alternative hypothesis: for more positive-going right-frontal ERP activity when preparing for episodic in comparison to non-episodic tasks (Herron & Wilding, 2004, 2006a). The number of participants and the *t* values for the main effects of interest, in the original and replication experiments are shown in **Table 10**.

Table 10. *t* values and sample sizes (*N*) from the original studies demonstrating a greater right-frontal positivity during preparation for episodic than for semantic memory retrieval, and for the replication attempt (Experiment 2).

Study	*t* value	*N*
Herron and Wilding (2004):		20
Main effect of cue-type (operation/semantic)	2.09	
Main effect of cue-type (location/semantic)	2.86	
Herron and Wilding (2006a):		16
Main effect of cue-type (location/semantic)	2.47	
Experiment 2:		32
Analysis strategy as 2004 study	1.21	
Analysis strategy as 2006a study	0.95	

As a replication of the episodic retrieval mode operation task effect in Herron and Wilding (2004) the $BF_{r0} = 0.87$ (**Figure 14A**). For a replication of the location effect the $BF_{r0} = 0.46$ (**Figure 14B**). These BFs indicate that the data provide anecdotal evidence in favour of the null hypothesis. For a replication of the location effect in Herron and Wilding (2006a) the $BF_{r0} = 0.33$ (**Figure 14C**). This indicates that the data provide substantial evidence for the null relative to the alternative hypothesis. Thus, the results of Experiment 2 suggest that the divergence reported on stay trials in the literature was not replicated here ($BF_{ave} = 0.55$, anecdotal evidence for the null hypothesis).

Herron and Wilding (2004)   **Herron and Wilding (2006a)**

A. Operation   **B.** Location   **C.** Location
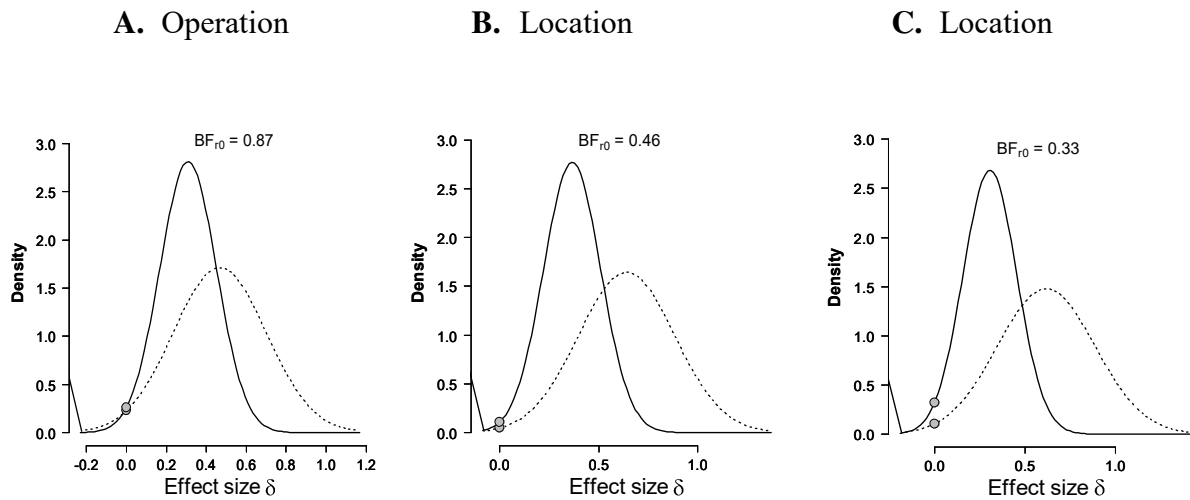
Figure 14. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the right-frontal positivity identified previously during preparation for episodic memory retrieval. In each panel, the dotted lines represent the posterior from the original study (Herron & Wilding, 2004, A & B; Herron & Wilding, 2006a, C), which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Experiment 2) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

4.5   Experiment 2: Discussion

Experiment 2 was conducted to test the proposal that the absence of the putative retrieval mode ERP index in Experiment 1 was due to matching the content of retrieval across the tasks. In addition, it was anticipated that the effects that were present in Experiment 1 also came about due to this attempt to equate the contents of retrieval. Thus, Experiment 2 was designed to permit an investigation of behaviour and associated ERP correlates when participants prepared to complete either an episodic memory task or a semantic memory task where the contents of the tasks were *unmatched*. This unmatched content design has been employed in the majority of previous studies demonstrating the proposed ERP index of retrieval mode (more positive-going activity following episodic task cues in comparison to baseline semantic task cues at right-frontal electrode sites for stay trials

90

only) (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). It was hypothesised that this effect would be replicated in this experiment.

There was, however, no evidence for this effect. In addition, BFs provided evidence in favour of the null hypothesis. These results suggest that there is not a generic retrieval mode ERP index that is identifiable when participants prepare for all episodic memory retrieval tasks. Regardless of whether content was matched across the tasks (Experiment 1) or whether it differed (Experiment 2), the pattern of activity identified previously in the literature was not evident. Perhaps it is more plausible that there are only indices of task- and content-specific retrieval orientations. In any case, why the neural signature previously identified was not evident in Experiments 1 or 2 requires further investigation, as well as consideration of design elements that may be important for explaining the differing findings across experiments.

In addition, in Experiment 2 there was no evidence of the ERP divergence seen in Experiment 1. One interpretation of the functional significance of the modulation in Experiment 1, in light of the results from Experiment 2, is that the activity reflects additional demands resulting from greater response competition in Experiment 1. On a subset of test trials in that experiment there was the potential for interference from different response mappings prompted by either the study history of an item or its semantic properties. For example, in the episodic task for some items an 'outside' response may have been required because of where they were presented at study, whereas they might in fact have been items most commonly found inside. In so far as interference from the task participants had just switched from carried over to the subsequent switch trial, it is possible that the preparatory modulations seen on switch trials in Experiment 1 are linked to the need to overcome this interference. For this account to work it is also necessary to assume that the way in which preparation for this challenge occurs is not identical when switching to the episodic or semantic task: if it was then no differences between the preparatory ERPs would be evident on switch trials. This account explains the absence of comparable effects in Experiment 2 because the potential for interference of this kind is diminished when content is unmatched across the episodic and semantic tasks.

Moreover, the presence of the source accuracy switch cost in Experiment 1 provides complementary support for this interpretation. In line with the argument developed above,

91

source retrieval may have been more demanding in Experiment 1, in comparison to Experiment 2, as a result of the level of response competition. Evidence from the wider task-switching literature suggests that response set overlap contributes towards age-differences in performance, particularly to the slower reaction times for older compared to younger participants, in task-switching experiments (Mayr, 2001). Thus, the potential additional demands resulting from different degrees of response set overlap may account for the switch cost for source accuracy in Experiment 1, as reconfiguration of the necessary task-set may have proved more difficult. Furthermore, in Experiment 2 there was no switch cost for reaction times specifically associated with the correct source judgments (see **Table 8**). Thus, on switch and stay trials, participants were equally efficient at getting the source of study location correct.

In previous content unmatched studies reaction time costs have consistently been observed (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002), whereas, additional costs for the accuracy of the memory judgments have not been identified reliably. Thus, in Experiment 2 behavioural reaction time switching costs were anticipated in the absence of performance accuracy switching costs (old/new discrimination accuracy/source accuracy), as obtained in the previous content unmatched studies. However, as well as reaction time switch costs, there was a switch cost for old/new discrimination. A possible explanation for this additional switch cost to recognition accuracy is the related nature of the task judgments in this experiment. Although the content was unmatched in Experiment 2, it was more matched than in the previous studies as here both task judgments involved spatial information. Working towards an understanding of why in some instances switch costs to accuracy are observed, and in some instances they are not, is of additional interest for further investigation, and a topic that will be returned to later in this thesis after consideration of further data points.

Turning to the question of the absence of effects in this experiment (and in Experiment 1) that resemble the putative index of retrieval mode, one possibility is that the index identified previously in the literature is a material-specific effect. It has been obtained only in studies where words have been employed at test (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). Pictures were employed in both Experiments 1 and 2. One way to test this possibility is to re-run Experiment 2 with the only change being the use of words as stimuli rather than pictures. Experiment 3 was designed in this way.

5. CHAPTER 5. Experiment 3: word stimuli & content unmatched

5.1 Experiment 3: Abstract

Experiment 3 was designed in order to investigate the possibility that the results obtained in Experiments 1 and 2 were due to the type of material processed at study and at test. In previous task-switching studies where ERPs were acquired, ERP indices of preparation for episodic memory retrieval comprised more positive-going activity at right-frontal scalp sites during preparation for the second trial of the episodic task. This effect has been identified as an index of retrieval mode (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). In these previous studies there were also no old/new discrimination switch costs reported. However, in Experiments 1 and 2 of this thesis there was no evidence for the retrieval mode ERP index, and there were significant old/new discrimination switch costs. It is plausible that these discrepancies are due to differences in the type of material to be retrieved. In this current experiment words were used as stimuli at both study and test, consistent with the previous studies in the literature (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). In addition, consistent with those studies, here the content to be retrieved was also unmatched across the tasks. However, the ERP activity related to each cue-type was only marginally significantly different, and inconsistent with patterns related to previous claims for a generic retrieval mode ERP index.

93

5.2   Experiment 3: Introduction

In an attempt to account for the absence of the ERP signature in Experiments 1 and 2, it was proposed that the index identified previously was a material-specific effect. In previous studies where the putative retrieval mode ERP index was identified, words were used as the stimuli. Thus, it is possible that the index identified previously was related to elements specific to verbal memory, and its absence in Experiments 1 and 2 was due to the use of pictures as stimuli. As Simons, Owen, et al. (2005) previously noted, findings from both functional imaging and neuropsychological studies of recognition memory have suggested differences in lateralisation depending on whether the stimuli were verbal/non-verbal. Furthermore, Lepage et al. (2000) noted that a potential caveat to their retrieval mode site meta-analysis was that the four studies for which the analysis was based on used almost entirely verbal stimuli. However, in the studies with non-verbal materials that were subsequently included in the meta-analysis there were some matches to the initial four-study verbal stimuli determined retrieval mode sites.

Wilckens et al. (2011) investigated preparatory processing for episodic memory retrieval using pictures instead of words, and ERPs for stay trials were more negative-going for the episodic task in comparison to the baseline semantic task. However, across experiment comparisons are difficult because a global vertex reference point was used instead of the average of the signal at the two mastoids that has been employed in the studies in this thesis, as well as in the majority of related published studies. In addition, the analysis reported by Wilckens et al. (2011) included three 'central superior' clusters of electrode sites along the midline only.

The design of this experiment was the same as Experiment 2 except that in this experiment words were used instead of pictures in all study and test blocks. It was anticipated that the ERP index previously associated with retrieval mode would be evident on stay trials at right-frontal scalp sites.

5.3   Experiment 3: Method

5.3.1   Participants

Data were collected from 33 participants based on the Bayesian Stopping Rule (Dienes, 2011) (see **General methods** & *Appendix A*), and the data from one participant were excluded due to excessive EEG artefact. Thus, the data from 32 participants (M_{age} = 20, range = 18-25, 27 female) were included in the analyses. Participants were paid £10 per hour.

5.3.2   Design

The stimuli were 240 words for the same objects selected from the International Picture Naming Project Database (Szekely et al., 2004) as in Experiments 1 and 2. The words were between three and ten letters in length, and the frequency range was between 0 and 7.396 CELEX log transformed (Szekely et al., 2004). All words were presented in Times New Roman font size 26. They were presented following the same protocol as in Experiments 1 and 2.

As in Experiment 2 there were approximately a third of the objects in each semantic category (88, 62, and 90, respectively), classified according to the size of the object depicted: smaller than a lunchbox, larger than a lunchbox but smaller than a suitcase, larger than both a lunchbox and a suitcase. For the semantic classification, the mean inter-rater reliability of three raters was 0.66. In this experiment there were ten study-test cycles instead of the five used in Experiment 2. Piloting indicated that this design yielded response accuracy that was equivalent to Experiment 2 (where the stimuli were pictures rather than words). For each cycle each of the stimuli from each semantic category were randomly assigned to one of ten lists. Thus, each list contained 24 objects: approximately nine small, six medium, and nine large objects. Again, two additional practice blocks were formed and used to familiarise participants with the experiment demands.

At study, 12 of the objects were either presented inside or outside an abstract outline of a building, and participants were required to indicate whether the object appeared inside or outside as in Experiments 1 and 2 (see **Figure 6**. Experiment 1. p. 66).

95

The test phase was identical to Experiment 2, with the exceptions that the stimuli were words, and there were only 24 items (12 old, 12 new) in each of ten test blocks.

5.3.3 Procedure

The sequence within each trial was as in Experiments 1 and 2 (see **Figure 6**. Experiment 1. p. 66), however, words were presented instead of pictures at both study and test. Trials on which responses were faster than 300ms or slower than 4000ms were counted as errors and excluded from the behavioural analyses (1.0% of the trials).

5.3.4 EEG acquisition

The EEG was recorded following the same protocol as in Experiments 1 and 2 (see **General methods**).

5.4 Experiment 3: Results

5.4.1 Frequentist statistics

On average, 84% of the switch and stay trials contributed to the ERP cue data in this experiment for each participant. The mean trial numbers contributing to the ERPs (ranges in parenthesis) were: episodic switch = 47 (26-53), episodic stay = 50 (28-60), semantic switch = 46 (28-55), semantic stay = 50 (29-59). ERPs elicited by the test items following the preparatory cues are included in *Appendix B*.

5.4.1.1 Behavioural analyses

During the study phases participants correctly responded 'inside' or 'outside' on 98% of trials. **Table 11** shows the response accuracy data for the test phases.

Table 11. Response accuracies for each task on switch and stay trials in Experiment 3. Standard deviations are in parentheses.

	Switch	Stay
Episodic task:		
Old/new discrimination (*Pr*)	0.65 (0.20)	0.69 (0.23)
P(correct source)	0.73 (0.19)	0.75 (0.15)
Correct rejection	0.83 (0.16)	0.84 (0.16)
Semantic task:		
Correct classification	0.79 (0.06)	0.76 (0.07)

Discrimination scores (discrimination index: $Pr = p$(hit) - p(false alarm), Snodgrass & Corwin, 1988) for the episodic task were above zero for both trial-types (switch: 0.65, stay: 0.69) and a two-tailed *t* test revealed a trend for greater on stay trials than on switch trials ($p = 0.091$; one-tailed: $t(31) = 1.74$, $p = 0.046$, $d_z = 0.31$, 62% CL).

The conditional probabilities of correct source judgments collapsed across the inside/outside dimension were reliably above chance in both cases (switch: $t(31) = 6.91$, $p < 0.001$, $d_z = 1.22$, 89% CL; stay: $t(31) = 9.20$, $p < 0.001$, $d_z = 1.63$, 95% CL). Source accuracies in the episodic task and the semantic task accuracies were not statistically different across trial-types.

A 2x2x2 ANOVA was conducted on the reaction times (see **Table 12**) for response accuracy categories (correct episodic source or semantic categorisation: when old/when new) separated according to task (episodic/semantic) and trial-type (switch/stay). Main effects of trial-type ($F(1, 31) = 15.23$, $p < 0.001$, $d_z = 0.69$, 75% CL), response accuracy category ($F(1, 31) = 62.63$, $p < 0.001$, $d_z = 1.40$, 92% CL), and task ($F(1, 31) = 4.42$, $p < 0.05$, $d_z = 0.37$, 64% CL), were moderated by interactions between task and trial-type ($F(1, 31) = 8.49$, $p < 0.05$, $\eta_p^2 = 0.21$) and task and response accuracy ($F(1, 31) = 28.80$, $p < 0.001$, $\eta_p^2 = 0.48$).

Looking at each task separately there were main effects of trial-type ($F(1, 31) = 16.23$, $p < 0.001$, $d_z = 0.71$, 76% CL) and response accuracy ($F(1, 31) = 54.39$, $p < 0.001$, $d_z = 1.30$, 90% CL) for the episodic task: reflecting slower responses for switch in comparison to stay, and for correct when old than correct when new responses. For the semantic task

there was also a main effect of response accuracy ($F(1, 31) = 11.73$, $p < 0.05$, $d_z = 0.60$, 73% CL) reflecting significantly slower correct when old than correct when new responses.

Table 12. Reaction times (ms) for each task on switch and stay trials in Experiment 3. Standard deviations are in parentheses.

	Switch	Stay
Episodic task:		
Correct source	1656 (342)	1551 (375)
Correct new	1332 (246)	1215 (245)
Semantic task:		
Correct old	1395 (316)	1413 (264)
Correct new	1359 (264)	1305 (236)

5.4.1.2 ERP analyses: cue data

As in the previous experiments the ERPs elicited by the two cues indicating which task to complete were analysed over an 800 to 1900ms time window, and the initial analysis included the same 12 sites distributed over fronto-central regions (F3/F4, F5/F6, F7/F8, C3/C4, C5/C6, T7/T8).

Figure 15 shows the grand-averaged ERP waveforms for each cue-type at right anterior and central sites, separated for switch and stay trials. Scalp maps depicting the differences between the scalp distributions of the ERPs associated with the cue-types are shown in **Figure 16**, again separated according to switch and stay trials (A & B).
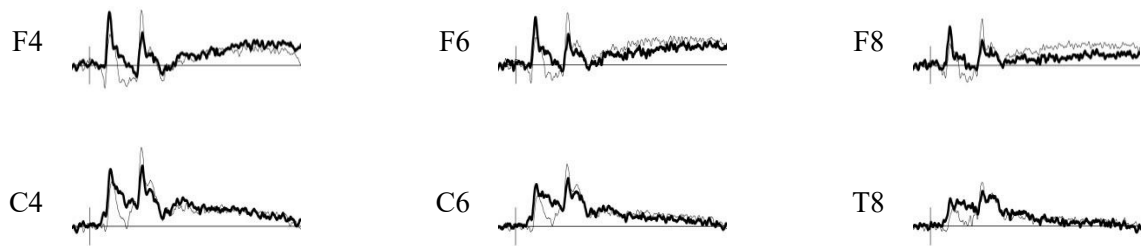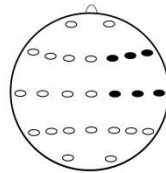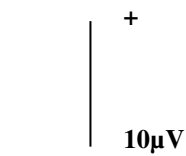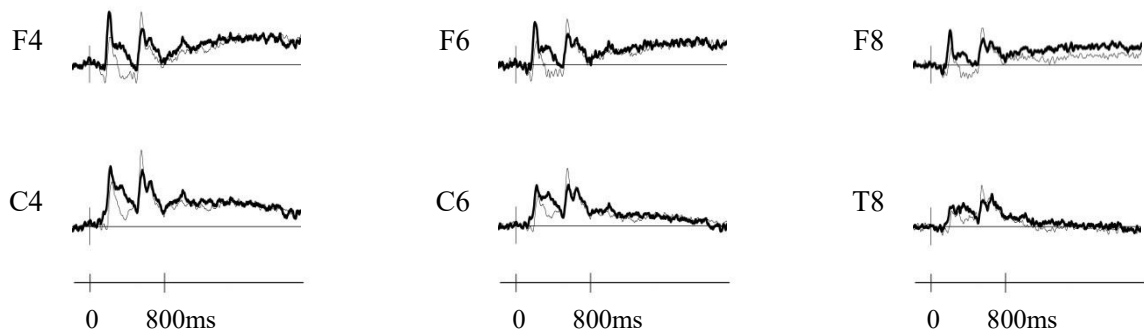
Switch

Stay

Figure 15. Grand averaged ERPs for Experiment 3 separated according to trial-type (switch/stay) and cue-type (episodic/semantic) for right anterior (F4, F6, F8) and right central electrode sites (C4, C6, T8).

A. Episodic Switch – Semantic Switch   **B.** Episodic Stay – Semantic Stay

800 ms - 1900 ms   800 ms - 1900 ms

-0.73 µV   0 µV   0.85 µV   -0.72 µV   1.64 µV
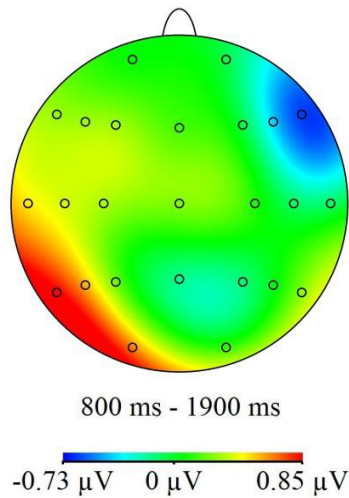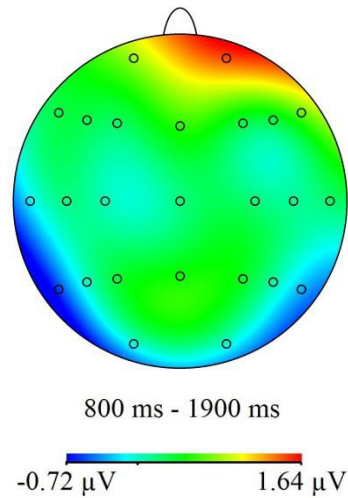
Figure 16. Topographic maps for Experiment 3: the differences between the scalp distributions of the ERPs associated with the episodic and semantic cues on switch (A) and stay (B) trials from 800 to 1900ms. The scale below each map denotes the voltage range (µV) of the differences between conditions.

The initial ANOVA was conducted incorporating the factors of cue-type (episodic/semantic), trial-type (switch/stay), location in the anterior-posterior plane (anterior/central), hemisphere (left/right), and site (inferior/mid-lateral/superior).

This revealed interactions involving the factors of cue-type, anterior-central dimension, and hemisphere ($F(1, 31) = 4.60$, $p < 0.05$, $\eta_p^2 = 0.13$), and cue-type, trial-type, and anterior-central dimension, and site ($F(1.8, 56.0) = 3.54$, $p < 0.05$, $\eta_p^2 = 0.10$). Separate ANOVAs were then carried out for switch and stay trials. These revealed trends for interactions between cue-type and hemisphere ($F(1, 31) = 3.56$, $p = 0.069$, $\eta_p^2 = 0.10$) and cue-type, anterior-central dimension and site ($F(1.9, 59.6) = 2.78$, $p = 0.072$, $\eta_p^2 = 0.08$) for switch trials only.

5.4.2 Bayesian statistics

5.4.2.1 Behavioural analyses: replication of accuracy switch costs

A Bayesian replication test (Dienes, 2011; Verhagen & Wagenmakers, 2014) was carried out in order to investigate whether the behavioural results from Experiment 3 provide evidence for a replication of the accuracy switch costs in Experiment 1. The number of participants and the *t* values for the main effects of interest, in the original experiment (1) and replication experiment (3), are shown in **Table 13**.

Table 13. *t* values and sample sizes (*N*) for the accuracy switch costs in Experiment 1, and for the replication attempt (Experiment 3).

Study	*t* value	*N*
Experiment 1:		24
Old/new discrimination switch cost	-2.56	
Source accuracy switch cost	-2.77	
Experiment 3:		32
Old/new discrimination switch cost	-1.74	
Source accuracy switch cost	-0.60	

For Experiment 3 as a replication of the old/new discrimination switch cost in Experiment 1 the $BF_{r0} = 2.09$ (**Figure 17A**). This BF indicates that the data provide anecdotal support for the alternative relative to the null hypothesis. For a replication of the source accuracy switch cost the $BF_{r0} = 0.2$ (**Figure 17B**). This BF indicates that the data provide substantial evidence for the null relative to the alternative hypotheses.

A. Old/new discrimination   **B.** Source
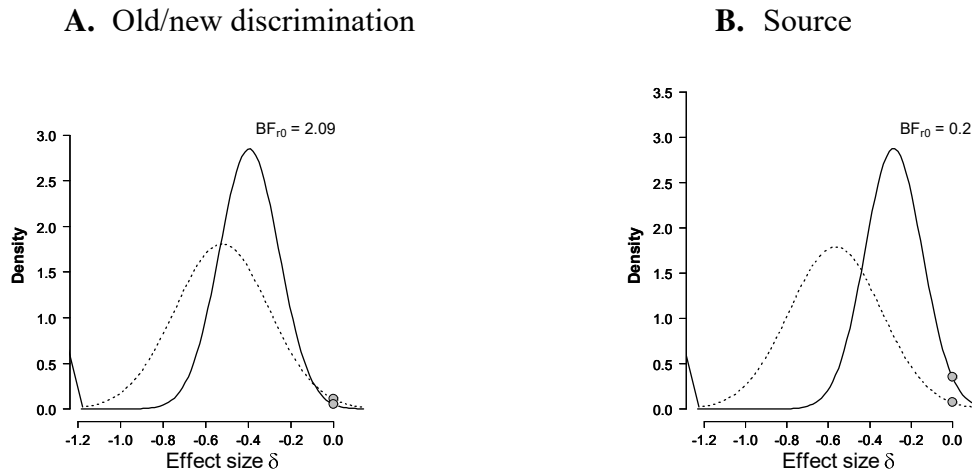
Figure 17. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the accuracy switch costs (A. Old/new discrimination; B. Source) in Experiment 1. In each panel, the dotted lines represent the posterior from the original study, which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Experiment 3) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

5.4.2.2 *ERP analyses: replication of right-frontal positivity during preparation for episodic memory retrieval*

A Bayesian replication test (Dienes, 2011; Verhagen & Wagenmakers, 2014) was carried out in order to investigate whether the ERP results from Experiment 3 provide support for the null or alternative hypothesis: for more positive-going right-frontal ERP activity when preparing for episodic in comparison to non-episodic tasks (Herron & Wilding, 2004, 2006a). The number of participants and the *t* values for the main effects of interest, in the original and replication experiments are shown in **Table 14**.

Table 14. *t* values and sample sizes (*N*) from the original studies demonstrating a greater right-frontal positivity during preparation for episodic than for semantic memory retrieval, and for the replication attempt (Experiment 3).

Study	*t* value	*N*
Herron and Wilding (2004):		20
Main effect of cue-type (operation/semantic)	2.09	
Main effect of cue-type (location/semantic)	2.86	
Herron and Wilding (2006a):		16
Main effect of cue-type (location/semantic)	2.47	
Experiment 3:		32
Analysis strategy as 2004 study	0.86	
Analysis strategy as 2006a study	0.81	

As a replication of the episodic retrieval mode operation task effect in Herron and Wilding (2004) the BF_{r0} = 0.5 (**Figure 18A**). This indicates that the data provide anecdotal evidence in favour of the null hypothesis. For a replication of the location effect the BF_{r0} = 0.23 (**Figure 18B**). For a replication of the location effect in Herron and Wilding (2006a) the BF_{r0} = 0.26 (**Figure 18C**). These BFs indicate that the data provide substantial evidence for the null relative to the alternative hypothesis. Thus, the results of Experiment 3 suggest that the divergence reported on stay trials in the literature was not replicated here (BF_{ave} = 0.33, substantial evidence for the null hypothesis).

Herron and Wilding (2004)　　　　　　　　　　　　**Herron and Wilding (2006a)**

A. Operation　　　　　　　　　**B.** Location　　　　　　　　　**C.** Location
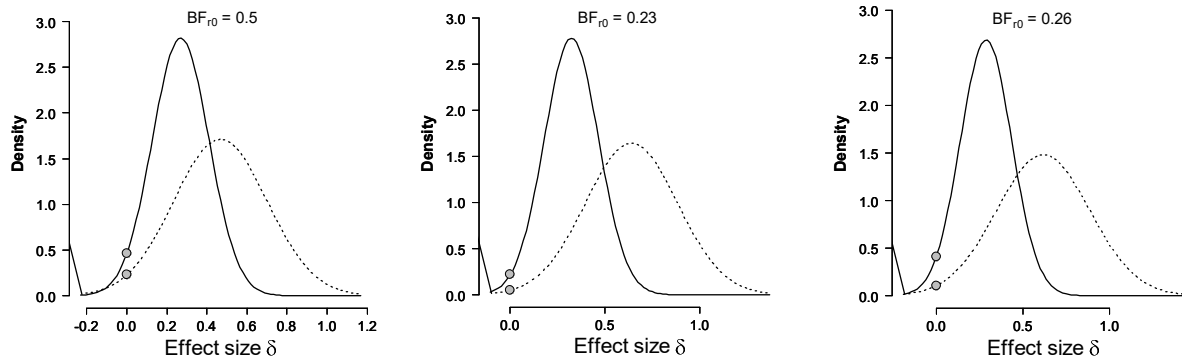
Figure 18. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the right-frontal positivity identified previously during preparation for episodic memory retrieval. In each panel, the dotted lines represent the posterior from the original study (Herron & Wilding, 2004, A & B; Herron & Wilding, 2006a, C), which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Experiment 3) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

5.5　Experiment 3: Discussion

Experiment 3 was conducted in order to assess the possibility that the absence of effects in Experiments 1 and 2 that have been reported previously was due to the use of pictures as stimuli. The design was the same as Experiment 2 except here words were used as the stimuli for all study and test blocks. The experiment was conducted in order to investigate whether: (i) preparatory ERP activity would differ for verbal material in ways reported previously, and (ii) the behavioural accuracy switch costs identified thus far in this thesis remained when words were the stimuli at study and at test.

There was no evidence for a switch cost for the accuracy of the source judgments, consistent with the results of Experiment 2. For old/new discrimination accuracy a two-tailed *t* test revealed a trend for greater discrimination scores on stay trials than on switch trials.

Moreover, the Bayesian analysis provided anecdotal evidence for the alternative hypothesis. Thus, here, unlike the previous studies investigating episodic memory retrieval of words, there was an indication of an effect of switching on the accuracy of the memory judgments (as evident in Experiments 1 and 2), in the absence of any evidence for the putative ERP index of retrieval mode. This failure to replicate what are apparently robust ERP effects reported previously in the literature in this experiment (and substantiated by the power analyses reports in this thesis) is arguably more striking than the replication failures in Experiments 1 and 2, because of the strong similarities between the design employed here and that employed in several previous studies (cf. Herron & Wilding, 2004; Herron & Wilding, 2006a; Morcom & Rugg, 2002). This outcome, in combination with the outcomes of Experiments 1 and 2, prompts careful consideration of other design elements that might be responsible for what comprise a set of somewhat disparate findings. Two elements are considered below: first, the predictability of the task sequence, and second, the trial timings.

In Experiments 1-3 the preparatory task cue sequence was predictable, and this may account for the differences between Experiment 3 and the previous studies where words were used as stimuli (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). In these previous studies either: (i) when the task switch would occur (Herron & Wilding, 2006a; Morcom & Rugg, 2002), or (ii) the task that participants were switching to (Herron & Wilding, 2004), was unpredictable. Thus, in Experiments 1-3 there is the possibility that participants had the opportunity to prepare for the subsequent task before the preparatory cue was presented, whereas this was not the case in other studies. If they became aware of the alternating runs switch-stay sequence, or if an automatic accommodation to the task structure occurred, preparation may have been set in train once the response to test items on preceding trials had been made, and this might explain the absence of effects reported previously, where preparation of this kind could not have been initiated.

Herron and Wilding (2006b) investigated the influence of the predictability of preparatory cue sequences on indices of preparation for episodic memory retrieval. The preparatory activity related to two episodic memory tasks was compared when there was a predictable and unpredictable cue sequence, and broadly similar indices of task-specific retrieval orientations were evident regardless of whether the sequence was predictable or not. However, an index was not evident in a further experiment where the preparatory cue-type was held constant within each cycle (a blocked design). Herron and Wilding (2006b)

105

concluded that the indices identified were associated with the initial adoption of a task-set, although, in this instance investigation was of different retrieval orientations and not retrieval mode (Herron & Wilding, 2006b).

Moving on to trial timings, in Experiments 1-3 there was a 1200ms period following the participant response before the next trial in the test phase. This was included in order to ensure that epochs for the response data were free from the preparatory cue for the next trial. In previous studies where electrophysiological indices of preparation for episodic memory retrieval have been investigated, the duration of the period after each response and before the onset of the next trial has varied. In Herron and Wilding (2004, 2006a) it was 500ms, and in Morcom and Rugg (2002) there was a fixed duration of 3200ms following the test word (during which participants responded). In these experiments, however, the task cue sequence was unpredictable or the task that participants were switching to was unpredictable, thus advanced preparation was not possible even if time was available. In Experiments 1-3, by contrast, participants may have prepared for the next trial during this period, and one way to test this is to reduce the length of this interval.

Herron and Wilding (2006a) investigated whether the length of the preparatory period *after* the preparatory cue influenced whether the ERP index of retrieval mode would be evident. When the period after the preparatory cue (the Cue-Target Interval, CTI; Meiran, Chorev, & Sapir, 2000) was lengthened to greater than 4000ms on the majority of trials, there was evidence of the proposed index of retrieval mode on stay trials which was sustained until the retrieval cue was presented (800-4000ms post-cue presentation). They concluded that time to prepare was not the primary determinant of the onset of retrieval mode (Herron & Wilding, 2006a). However, it remains plausible that preparation time *before* the preparatory cue has a role to play, particularly when there is a predictable task cue sequence.

The trial period in question here has also been described as the Response-Cue Interval (RCI, Meiran et al., 2000), and in behavioural studies it has been found that task-switching costs were reduced by increasing the RCI (Meiran et al., 2000). Hence there are additional forms of support for the view that this element of the task design is worthy of attention. In light of these considerations Experiment 4 had an unpredictable preparatory task cue sequence and the RCI was 500ms.

6. CHAPTER 6. Experiment 4: 500ms RCI & unpredictable cue sequence

6.1 Experiment 4: Abstract

In several ERP studies of preparatory retrieval processing, neural activity associated with preparing for episodic retrieval is more positive-going at right-frontal sites than activity associated with preparing to complete tasks with no episodic demand (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). This activity has been proposed to index retrieval mode and has most often been observed on stay trials in tasks requiring frequent task-switches. In contrast to the general task-switching literature (e.g. Meiran et al., 2000; Monsell et al., 2003; Rogers & Monsell, 1995), limited attention has been paid in ERP memory studies to within-trial design elements which might influence the opportunity people have to prepare for episodic retrieval. Here, the factors of predictability of task-switching requirements and time to prepare were altered relative to the parameters employed in the previous studies reported in this thesis. These alterations were made after consideration of design elements that might explain the null results obtained in Experiments 1-3.

When the sequence of the cue that signalled which task to complete was predictable, and the interval between participant response and the next cue (Response-Cue Interval, RCI) was 1200ms, there was no evidence for an ERP index that has been linked previously to retrieval mode (Experiments 1-3). An index was evident, however, in this experiment where the task cue sequence was unpredictable and the RCI was 500ms. Moreover, more positive-going activity was evident following the episodic task cue (in comparison to the semantic task cue) on the first trial of the task only (switch trials), contrasting with several previous findings where it was evident only on the second successive trial of the same task (stay trials). These findings highlight the need to consider design factors to constrain explanations for when preparation for episodic retrieval is evident, and, consequently, to understand the benefits that preparation for episodic retrieval afford.

6.2   Experiment 4: Introduction

In the experiments reported thus far in this thesis there was no indication of the preparatory ERP index linked to retrieval mode in several earlier studies. In order to account for the absence of the previously robust observation of a retrieval mode ERP index, various factors were considered systematically across Experiments 1-3. In summary, the index was not evident when the content was matched across the tasks (Experiment 1), for an unmatched content design with pictures (Experiment 2), or an unmatched content design with words as stimuli (Experiment 3).

Two elements to which some attention has been paid are the predictability of the task cue sequence (Herron & Wilding, 2006b), and the preparation time available (Herron & Wilding, 2006a). Herron and Wilding (2006a) lengthened the time following each task cue, over what had typically been used, and the putative index of retrieval mode remained evident on stay trials (and it was sustained until the test word was presented 4000ms later). This led the authors to conclude that time to prepare after a preparatory cue was not the primary determinant for the time course of this process. In addition, in an experiment in which preparation for different episodic memory tasks was investigated, indices of separate retrieval orientations were evident when the task cue sequence was predictable as well as when it was not (Herron & Wilding, 2006b).

However, in light of the results from Experiments 1-3, closer inspection of the parameters employed in the previous studies where electrophysiological indices of preparation for episodic memory retrieval have been investigated enabled further hypothesis generation. Collectively, the similarities and the differences in parameters set across the previous experiments, and the experiments reported in this thesis, provided an indication that the factors of predictability of the task cue sequence and preparation time may have influenced the results that were obtained.

To summarise briefly, in Herron and Wilding (2006a) the task cue sequence was unpredictable as four additional catch trials were inserted into each test block. In addition, the interval between participant response and the next cue (RCI) was 500ms. In Herron and Wilding (2004) the RCI was also 500ms, although, there were no catch trials included. However, the task cue sequence was again unpredictable. This was because participants were

required to switch between three tasks rather than two. Thus, in Herron and Wilding (2004), the task switch was predictable (as there were always two trials of each cue-type) but the task that participants were switching to was unpredictable (one of two alternatives). In Morcom and Rugg (2002), the RCI was longer than in the other previous experiments (3200ms, including response time). However, the task cue sequence was unpredictable: with switch, stay and stay+1 trials varying randomly. Thus, there was potentially time to prepare, however, the participants were unable to as the task switch was not predictable.

This is in contrast to Experiments 1-3 of this thesis, where the RCI was 1200ms and the task cue sequence was predictable. It was possible that in these experiments, participants were able to prepare for the task switch in advance of the preparatory task cue signalling the switch.

Experiment 4 was designed so that the preparation time (RCI) and the predictability of the preparatory task cue sequence differed from that employed in previous studies in this thesis. The matched content design from Experiment 1 of this thesis was adapted. In Experiment 1 the task cue sequence was predictable, and the RCI was 1200ms. In the current experiment, the task cue sequence was unpredictable and the RCI was 500ms. It was predicted that these design changes would result in the reinstatement of the putative retrieval mode index reported previously.

6.3   Experiment 4: Method

6.3.1   Participants

Data were collected from 35 participants based on the Bayesian Stopping Rule (Dienes, 2011) (see **General methods** & *Appendix A*), and the data from three participants were excluded due to excessive EEG artefacts. Thus, the data from 32 participants (M_{age} = 23, range = 18-30, 23 female) were included in the analyses. Participants were paid £10 per hour.

6.3.2   Design

The stimuli were the same 240 black line drawings of objects (Szekely et al., 2004) used in Experiments 1 and 2. The objects were presented on a monitor following the same protocol as in Experiments 1-3, and, as in Experiments 1 and 2, there were five study-test

cycles. The 240 objects were classified in the same three semantic categories as in Experiment 1 (according to where they were commonly found: inside, outside or both), and 80 stimuli from each semantic category were assigned to one of the five lists. Thus, each list contained 48 objects with 16 from each semantic category (inside/outside/both). Two additional practice blocks, half the length of the other five study-test blocks, were formed and used to familiarise participants with the experiment demands.

All aspects of the study phase were identical to Experiment 1 (see **Figure 6**. Experiment 1. p. 66). At test, the only departure from the design of Experiment 1 was in the trial sequence. In this experiment, rather than presenting the same cue-type for two consecutive trials, the sequence of the cue that signalled which task to complete was unpredictable. In order to implement this, eight trials from each test block were used as 'catch' (stay+1, or stay+2) trials. Thus, for each test block (for each task) there was one four-trial run, two three-trial runs, five two-trial runs, and four one-trial runs. The order of these was randomised within each test block. This meant that within each of the five test phases there were 12 'switch' trials, and 12 'repeat' trials (eight 'stay', three 'stay+1', one 'stay+2') for each task.

Item status was counterbalanced across participants in the same way as in Experiment 1, with 'switch' and 'repeat' trials counterbalanced accordingly. During the test phase responses were made using the same fingers as at study, with the addition of the index finger of the other hand to indicate 'new' or 'both', for the episodic and the semantic task respectively. The hands used for the judgments were counterbalanced across participants.

6.3.3 Procedure

The procedure for the study phase followed the same trial sequence as in Experiments 1-3. However, the procedure for the test phase differed in that here the interval between participant response and the next cue (RCI) was 500ms (rather than 1200ms; see **Figure 6**. Experiment 1. p. 66, & **Figure 19**). Trials on which responses were faster than 300ms or slower than 4000ms were counted as errors and excluded from the behavioural analyses (0.9% of the trials).
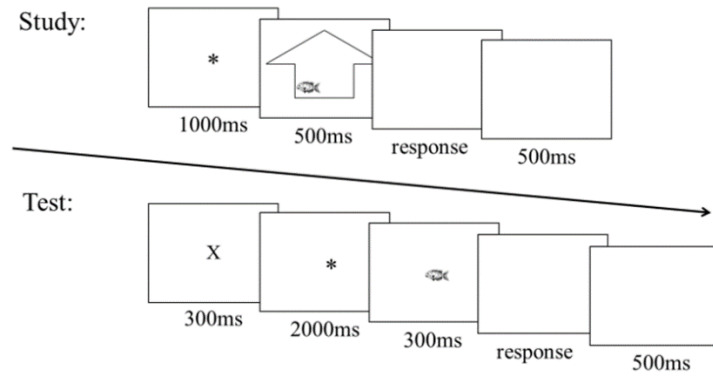
Figure 19. Schematic illustration of the trial sequence and timings at study and test for Experiment 4.

6.3.4 EEG acquisition

The EEG was recorded following the same protocol as in Experiments 1-3 (see **General methods**).

6.4 Experiment 4: Results

6.4.1 Frequentist statistics

On average, 84% of the switch and stay trials contributed to the ERP cue data in this experiment for each participant. The mean trial numbers contributing to the ERPs (ranges in parentheses) were: episodic switch = 50 (31-56), episodic stay = 34 (21-39), semantic switch = 49 (32-55), semantic stay = 34 (23-40). ERPs elicited by the test items following the preparatory cues are included in ***Appendix B***.

6.4.1.1 Behavioural analyses

During the study phases participants correctly responded 'inside' or 'outside' on 97% of trials. **Table 15** shows the response accuracy data for the test phases.

Table 15. Response accuracies for each task on switch and stay trials in Experiment 4. Standard deviations are in parentheses.

	Switch	Stay
Episodic task:		
Old/new discrimination (*Pr*)	0.68 (0.14)	0.72 (0.16)
P(correct source)	0.76 (0.13)	0.83 (0.12)
Correct rejection	0.79 (0.13)	0.84 (0.12)
Semantic task:		
Correct classification	0.75 (0.08)	0.74 (0.10)

Discrimination scores (discrimination index: $Pr = p$(hit) - p(false alarm), Snodgrass & Corwin, 1988) for the episodic task were above zero for both trial-types (switch: 0.68, stay: 0.72) and a two-tailed t test revealed a trend for greater on stay trials than on switch trials ($p = 0.069$; one-tailed: $t(31) = 1.89$, $p = 0.035$, $d_z = 0.33$, 63% CL).

The conditional probabilities of correct source judgments collapsed across the inside/outside dimension were reliably above chance in both cases (switch: $t(31) = 11.84$, $p < 0.001$, $d_z = 2.09$, 98% CL; stay: $t(31) = 15.40$, $p < 0.001$, $d_z = 2.72$, 99.7% CL). Performance was superior on stay trials ($t(31) = 3.56$, $p = 0.001$, $d_z = 0.63$, 74% CL). The semantic task accuracies were not statistically different across trial-types.

A 2x2x2 ANOVA was conducted on the reaction times (see **Table 16**) for response accuracy categories (correct episodic source or semantic categorisation: when old/when new) separated according to task (episodic/semantic) and trial-type (switch/stay). There was a main effect of trial-type ($F(1, 31) = 16.81$, $p < 0.001$, $d_z = 0.72$, 77% CL) reflecting slower responses on switch than stay trials. There were also main effects of response accuracy category ($F(1, 31) = 30.04$, $p < 0.001$, $d_z = 0.97$, 83% CL) and task ($F(1, 31) = 15.12$, $p < 0.001$, $d_z = 0.69$, 75% CL) moderated by an interaction between these factors ($F(1, 31) = 8.57$, $p < 0.05$, $\eta_p^2 = 0.22$). This reflects significantly slower correct when old than correct when new responses for both tasks, and slower responses for the semantic task in comparison to the episodic task for correct new responses ($p < 0.001$).

Table 16. Reaction times (ms) for each task on switch and stay trials in Experiment 4. Standard deviations are in parentheses.

	Switch	Stay
Episodic task:		
Correct source	1521 (340)	1441 (357)
Correct new	1312 (242)	1236 (232)
Semantic task:		
Correct old	1573 (321)	1480 (275)
Correct new	1500 (253)	1389 (232)

6.4.1.2 ERP analyses: cue data

As in the previous experiments the ERPs elicited by the two cues indicating which task to complete were analysed over an 800 to 1900ms time window, and the initial analysis included the same 12 sites distributed over fronto-central regions (F3/F4, F5/F6, F7/F8, C3/C4, C5/C6, T7/T8).

Figure 20 shows the grand-averaged ERP waveforms for each cue-type at right anterior and central sites, separated for switch and stay trials. Scalp maps depicting the differences between the scalp distributions of the ERPs associated with the cue-types are shown in **Figure 21**, again separated according to switch and stay trials (A & B). These figures demonstrate a greater relative positivity at right-frontal sites for the episodic task, on switch trials.

113
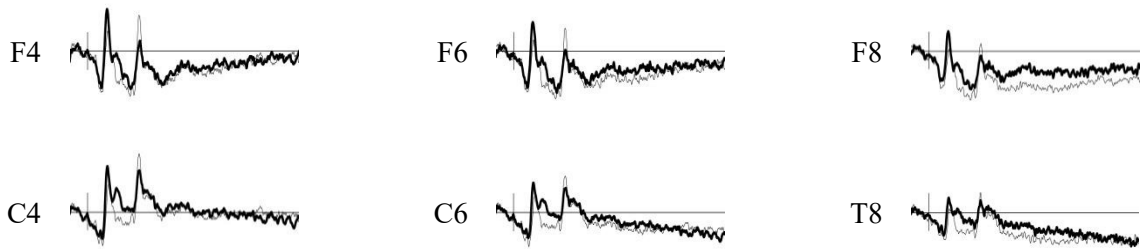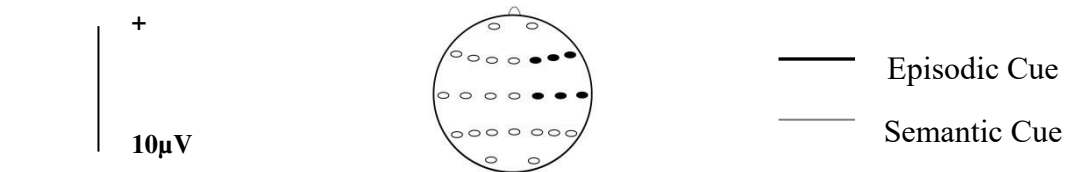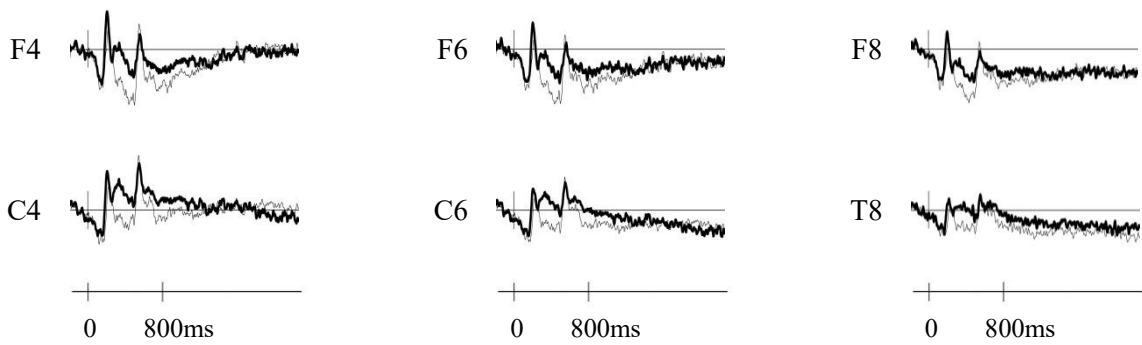
Switch

Stay

Figure 20. Grand averaged ERPs for Experiment 4 separated according to trial-type (switch/stay) and cue-type (episodic/semantic) for right anterior (F4, F6, F8) and right central electrode sites (C4, C6, T8).

A. Episodic Switch – Semantic Switch   **B.** Episodic Stay – Semantic Stay

800 ms - 1900 ms

-0.72 µV 0 µV   1.14 µV

800 ms - 1900 ms

-0.27 µV   1.00 µV
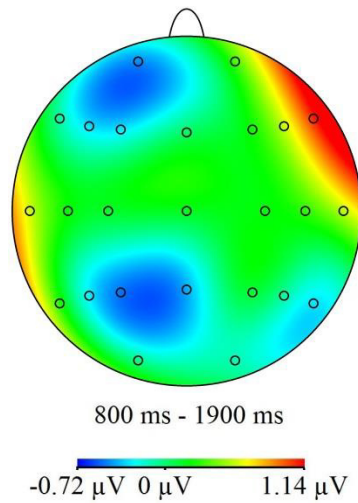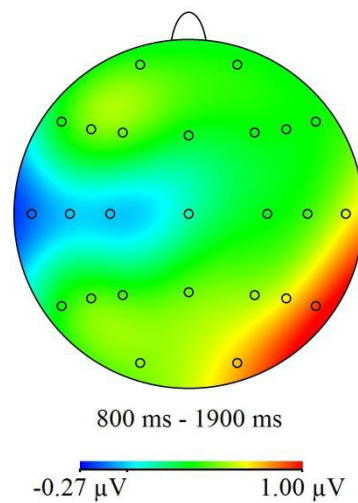
Figure 21. Topographic maps for Experiment 4: the differences between the scalp distributions of the ERPs associated with the episodic and semantic cues on switch (A) and stay (B) trials from 800 to 1900ms. The scale below each map denotes the voltage range (µV) of the differences between conditions.

The initial ANOVA was conducted incorporating the factors of cue-type (episodic/semantic), trial-type (switch/stay), location in the anterior-posterior plane (anterior/central), hemisphere (left/right), and site (inferior/mid-lateral/superior).

This revealed an interaction between cue-type, trial-type, anterior-central dimension, and hemisphere ($F(1, 31) = 18.04$, $p < 0.001$, $\eta_p^2 = 0.37$) (and a trend between cue-type and hemisphere: $F(1, 31) = 4.18$, $p = 0.05$, $\eta_p^2 = 0.12$). Separate ANOVAs were then carried out for switch and stay trials. These revealed a cue-type by anterior-central by hemisphere interaction for both trial-types (switch: $F(1,31) = 10.03$, $p < 0.05$, $\eta_p^2 = 0.24$; stay: $F(1,31) = 9.72$, $p < 0.05$, $\eta_p^2 = 0.24$).

Follow up ANOVAs were subsequently carried out for the anterior and central sites. For switch trials at the anterior sites, there was an interaction between cue-type and hemisphere ($F(1, 31) = 8.94$, $p < 0.05$, $\eta_p^2 = 0.22$), which reflects more-positive going activity for the episodic task in comparison to the semantic task over right hemisphere sites, and more positive-going activity for the semantic task than the episodic task over left hemisphere sites. At central sites there were no significant outcomes involving the factor of cue-type.

115

For stay trials at the anterior sites, there were no significant outcomes involving the factor of cue-type. However, at central sites there was an interaction between cue-type and hemisphere ($F(1,31) = 4.62$, $p < 0.05$, $\eta_p^2 = 0.13$). This cue-type by hemisphere interaction also reflects more-positive going activity for the episodic task in comparison to the semantic task over right hemisphere sites, and more positive-going activity for the semantic task than the episodic task over left hemisphere sites.

6.4.2 Bayesian statistics

6.4.2.1 Behavioural analyses: replication of accuracy switch costs

A Bayesian replication test (Dienes, 2011; Verhagen & Wagenmakers, 2014) was carried out in order to investigate whether the behavioural results from Experiment 4 provide evidence for a replication of the accuracy switch costs in Experiment 1. The number of participants and the *t* values for the main effects of interest, in the original experiment (1) and replication experiment (4), are shown in **Table 17**.

Table 17. *t* values and sample sizes (*N*) for the accuracy switch costs in Experiment 1, and for the replication attempt (Experiment 4).

Study	*t* value	*N*
Experiment 1:		24
Old/new discrimination switch cost	-2.56	
Source accuracy switch cost	-2.77	
Experiment 4:		32
Old/new discrimination switch cost	-1.89	
Source accuracy switch cost	-3.56	

For Experiment 4 as a replication of the old/new discrimination switch cost in Experiment 1 the $BF_{r0} = 2.87$ (**Figure 22A**). For a replication of the source accuracy switch cost the $BF_{r0} = 140.41$ (**Figure 22B**). These BFs indicate that the data provide anecdotal and decisive evidence, respectively, for the alternative relative to the null hypotheses.
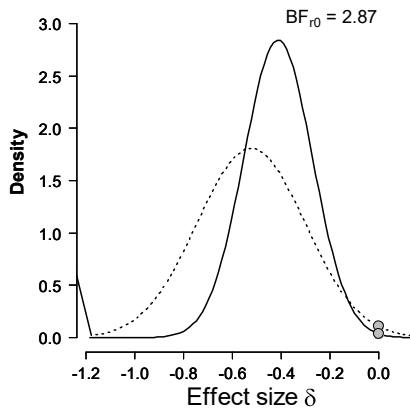
A. Old/new discrimination   **B.** Source
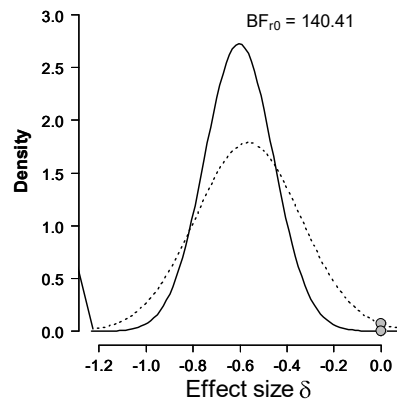
Figure 22. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the accuracy switch costs (A. Old/new discrimination; B. Source) in Experiment 1. In each panel, the dotted lines represent the posterior from the original study, which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Experiment 4) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

6.4.2.2   ERP analyses: replication of right-frontal positivity during preparation for episodic memory retrieval

A Bayesian replication test (Dienes, 2011; Verhagen & Wagenmakers, 2014) was carried out in order to investigate whether the ERP results from Experiment 4 provide support for the null or alternative hypothesis: for more positive-going right-frontal ERP activity when preparing for episodic in comparison to non-episodic tasks (Herron & Wilding, 2004, 2006a). The number of participants and the *t* values for the main effects of interest, in the original and replication experiments are shown in **Table 18**.

117

Table 18. *t* values and sample sizes (*N*) from the original studies demonstrating a greater right-frontal positivity during preparation for episodic than for semantic memory retrieval, and for the replication attempt (Experiment 4).

Study	*t* value	*N*
Herron and Wilding (2004):		20
Main effect of cue-type (operation/semantic)	2.09	
Main effect of cue-type (location/semantic)	2.86	
Herron and Wilding (2006a):		16
Main effect of cue-type (location/semantic)	2.47	
Experiment 4:		32
Analysis strategy as 2004 study	1.06	
Analysis strategy as 2006a study	0.49	

As a replication of the episodic retrieval mode operation task effect in Herron and Wilding (2004) the $BF_{r0} = 0.68$ (**Figure 23A**). For a replication of the location effect the $BF_{r0} = 0.34$ (**Figure 23B**). These BFs indicate that the data provide anecdotal evidence in favour of the null hypothesis. For a replication of the location effect in Herron and Wilding (2006a) the $BF_{r0} = 0.16$ (**Figure 23C**). This indicates that the data provide substantial evidence for the null relative to the alternative hypothesis. Thus, the results of Experiment 4 suggest that the divergence reported on stay trials in the literature was not replicated here ($BF_{ave} = 0.39$, anecdotal evidence for the null hypothesis).

A. Operation   **B.** Location   **C.** Location
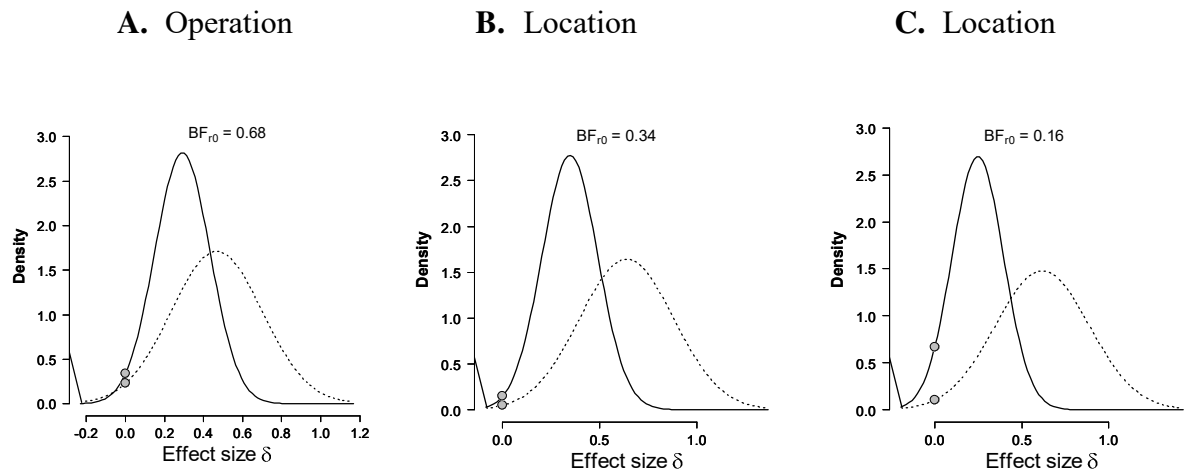
Figure 23. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the right-frontal positivity identified previously during preparation for episodic memory retrieval. In each panel, the dotted lines represent the posterior from the original study (Herron & Wilding, 2004, A & B; Herron & Wilding, 2006a, C), which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Experiment 4) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

6.5   Experiment 4: Discussion

This experiment was designed to investigate the possibility that the predictability of the preparatory cue sequence and the RCI were determinants of the conditions under which a signal linked previously with retrieval mode would be evident.   The design employed here was the same as Experiment 1, with matched content across the tasks and pictures as stimuli, but here the task cue sequence was unpredictable and the RCI was shortened to 500ms (1200ms in Experiment 1).

In keeping with the findings in Experiment 1, there was a divergence in the preparatory activity associated with each cue-type on switch trials. However, the divergence was not equivalent to that observed in Experiment 1. Rather, there was more positive-going activity following the episodic task cue in comparison to the semantic task cue over right-anterior sites, mirroring the findings obtained in previous studies.

In contrast to the divergence reported in the previous literature, however, the more positive-going right-frontal activity following the episodic task cue was identified on switch trials rather than on stay trials. A potential explanation for this difference is that in the current experiment the contents were matched across the episodic and semantic tasks. This might have decreased the degree of cognitive reconfiguration required when switching between these tasks, relative to the reconfiguration challenges in previous studies when content was not matched. A similar explanation was put forward to explain why indices of retrieval orientation emerge on switch as opposed to stay trials (Herron & Wilding, 2006b; J. D. Johnson & Rugg, 2006). The researchers argued that less reconfiguration is required when switching between two episodic than between one episodic and one semantic task as there is a greater degree of overlap between the relevant cognitive operations (Herron & Wilding, 2006b; J. D. Johnson & Rugg, 2006).

Similarly, the related nature of the task judgments may also explain why there was a trend (two-tailed) for a switch cost for old/new discrimination accuracy in this experiment. Again, here the Bayesian analysis provided anecdotal evidence for the alternative hypothesis. An indication of this switching cost has been observed consistently across Experiments 1-4 of this thesis. In Experiments 2 and 3 the content was unmatched across the tasks; however, in all four experiments the judgments involved spatial information. Thus, the more related nature of the task judgments may account for these additional accuracy costs. In this experiment there was also a switch cost for the source accuracy of the memory judgments, as obtained in Experiment 1 where the contents of the judgments were also matched across the tasks. This finding is consistent with the earlier claim that this effect potentially reflects interference resulting from response competition across the episodic and semantic tasks.

In summary, these findings highlight the need to consider procedural design factors to constrain explanations for when there is electrophysiological evidence for preparation for episodic retrieval. Experiments 1-4 demonstrate that the presence and timing of the index is not consistent across different task demands. A more in-depth discussion of the implications of the ERP results from Experiments 1-4 is included in the **General discussion** for this thesis. In the final chapter of this thesis containing new empirical data, the outcomes of two behavioural experiments are reported. They were designed to permit an investigation of the boundary conditions for when switch costs occur in tasks requiring episodic memory judgments.

Part 2. Frequency of task-switches and behavioural performance measures for episodic memory retrieval

7. CHAPTER 7. Experiment 5: stay+1 trials, & Experiment 6: alternating costs

7.1 *Experiments 5 & 6: Abstract*

Two additional experiments were designed to investigate the behavioural costs that are evident when switching between memory tasks. The first of these (Experiment 5) was designed to investigate the time course of the costs identified previously. Additional 'stay+1' trials were included to determine whether costs extended beyond switch trials. The second (Experiment 6) was designed to investigate whether the switch costs identified in earlier experiments reflect in part an 'alternating cost' as a result of alternating frequently between tasks. This was investigated by comparing reaction times and accuracy measures in 'blocked' phases with no switching requirements, and 'alternating' phases in which switches were required frequently. Switch costs for accuracy were restricted to switch trials in Experiment 5; however, there were reaction time costs on stay trials relative to stay+1 trials. In Experiment 6 reaction times were slower on stay trials in alternating phases than the average reaction time in blocked phases. These outcomes provide a starting point for understanding boundary conditions for performance costs which arise when participants switch in and out of completing an episodic memory task.

7.2   Experiments 5 & 6: Introduction

In most laboratory studies of memory many trials are completed in succession where the same kind of memory judgment is made. This does not mirror how we typically use our memories in everyday life, where memory retrieval is something we do interspersed and often in parallel with other tasks. When asked to switch between episodic and semantic memory tasks, participants are faster when the same task is completed again in comparison to when they switch to another task (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). Moreover, in one instance, Herron and Wilding (2006a) noted a switch cost for the accuracy of source judgments.

Experiments 1-4 of this thesis enabled further investigation of reaction time and accuracy costs when task-switches are required. In all four experiments, reaction time costs were evident as well as indications of old/new discrimination costs. In Experiments 1 and 4 there were also accuracy costs for source memory judgments. Considerations of the reasons for apparent inconsistences across studies will be covered in subsequent discussions. Of immediate interest here are the boundary conditions for when switch costs are revealed in memory switching paradigms.

7.3   Experiment 5: Introduction

Experiment 5 was conducted to investigate whether there were any further significant improvements in performance for *stay+1* trials relative to stay trials. The design was the same as Experiment 1, except that here three trials of the same task were completed in a row. It was anticipated that the switch costs evident in Experiments 1 and 4 would be replicated. Using a longer run length of task repetitions enabled investigation of any additional performance changes.

In general switching tasks not requiring memory judgments, there is mixed evidence for a gradual approach to asymptotic performance (Mayr, 2001; Rogers & Monsell, 1995; Salthouse, Fristoe, McGuthry, & Hambrick, 1998). In Herron and Wilding (2006a), where stay+1 trials were included and memory judgments were required, there was some evidence that reaction times and source accuracy improved from stay to stay+1 trials.

In this experiment, significant improvements for reaction time and accuracy for stay+1 trials would necessitate further investigation of the time course of switch costs. On the other hand, the absence of switch costs on stay trials would go some way to establishing boundary conditions for these costs.

7.4   Experiment 5: Method

7.4.1   Participants

A sample size of 24 participants was decided *a priori* based on power analyses (see ***Appendix A***) and counterbalancing considerations. Data were collected from 27 participants, and the data from three participants were excluded due to old/new discrimination scores falling below 0.10. Thus, the data from 24 participants (M_{age} = 20, range = 18-24, 23 female) were included in the analyses. In this experiment, participants were paid £6 per hour.

7.4.2   Design

The stimuli were 384 black line drawings of objects selected from the International Picture Naming Project Database (Szekely et al., 2004). The corresponding name for each object was between three and ten letters in length, the percentage picture naming frequency was above 0.60, and the frequency range was between 0 and 7.396 CELEX log transformed (Szekely et al., 2004). The objects were presented following the same protocol as in Experiments 1-4.

As in Experiment 1 the objects were classified into one of three semantic categories, according to where they were usually found: inside, outside or both. There were 128 objects in each semantic category. For the semantic classification, the mean inter-rater reliability of three raters was 0.70. The experiment comprised eight study-test cycles, and the 128 stimuli from each semantic category were randomly assigned to one of eight lists. Thus, each list contained 48 objects: 16 from each semantic category (inside/outside/both). Two additional practice blocks, half the length of the other eight study-test blocks, were formed and used to familiarise participants with the experiment demands.

All aspects of the study phases were identical to Experiment 1 (see **Figure 6**. Experiment 1. p. 66). At test, the only departure from the design of Experiment 1 was in the trial sequence. Each cue-type was always presented for three consecutive trials: switch, stay,

and stay+1 trials. Trials where the cue signalled the same task as on the previous two trials are referred to as *stay+1* trials. Whether the object appeared on a switch, stay, or stay+1 trial, the task status (episodic/semantic), and the old/new status of the object were counterbalanced across participants. As in Experiment 1, during the test phase responses were made using the same fingers as at study, with the addition of the index finger of the other hand to indicate 'new' or 'both', for the episodic and the semantic task respectively. The hands used for the judgments were counterbalanced across participants.

7.4.3 Procedure

The sequence within each trial was as in Experiments 1-3 (see **Figure 6**. Experiment 1. p. 66). Trials on which responses were faster than 300ms or slower than 4000ms were counted as errors and excluded from analyses (2.1% of the trials).

7.5 Experiment 5: Results

7.5.1 Frequentist statistics

7.5.1.1 Behavioural analyses

During the study phases participants correctly responded 'inside' or 'outside' on 96% of trials. **Table 19** shows the response accuracy data for the test phases.

Table 19. Response accuracies for each task on switch, stay and stay+1 trials in Experiment 5. Standard deviations are in parentheses.

	Switch	Stay	Stay+1
Episodic task:			
Old/new discrimination (*Pr*)	0.62 (0.24)	0.69 (0.20)	0.66 (0.23)
P(correct source)	0.73 (0.14)	0.74 (0.13)	0.76 (0.16)
Correct rejection	0.77 (0.16)	0.83 (0.14)	0.83 (0.14)
Semantic task:			
Correct classification	0.67 (0.10)	0.67 (0.10)	0.70 (0.09)

Discrimination scores (discrimination index: $Pr = p(\text{hit}) - p(\text{false alarm})$, Snodgrass & Corwin, 1988) for the episodic task were above zero for all trial-types (switch: 0.62, stay: 0.69, stay+1: 0.66) and differed significantly between switch and stay trials only ($t(23) = 2.45$, $p < 0.05$, $d_z = 0.50$, 69% CL).

The conditional probabilities of correct source judgments collapsed across the inside/outside dimension were reliably above chance in all three cases (switch: $t(23) = 7.80$, $p < 0.001$, $d_z = 1.59$, 94% CL; stay: $t(23) = 8.80$, $p < 0.001$, $d_z = 1.80$, 96% CL; stay+1: $t(23) = 8.26$, $p < 0.001$, $d_z = 1.69$, 95% CL). Source accuracies were not statistically different across the trial-types. For the semantic task, paired t tests revealed a significant difference between performance accuracies for switch and stay+1 trials only ($t(23) = 2.37$, $p < 0.05$, $d_z = 0.48$, 69% CL).

A 2x2x3 ANOVA was conducted on the reaction times (see **Table 20**) for response accuracy categories (correct episodic source or semantic categorisation: when old/when new) separated according to task (episodic/semantic) and trial-type (switch/stay/stay+1). A main effect of trial-type ($F(1.94, 44.66) = 27.97$, $p < 0.001$, $\eta_p^2 = 0.55$) reflects significantly slower responses for switch than stay ($t(23) = 3.69$, $p = 0.001$, $d_z = 0.75$, 77% CL), switch than stay+1 ($t(23) = 7.88$, $p < 0.001$, $d_z = 1.61$, 95% CL), and stay than stay+1 trials ($t(23) = 3.62$, $p = 0.001$, $d_z = 0.74$, 77% CL).

There were also main effects of response accuracy category ($F(1, 23) = 36.42$, $p < 0.001$, $d_z = 1.23$, 89% CL) and task ($F(1, 23) = 14.88$, $p = 0.001$, $d_z = 0.79$, 78% CL), moderated by an interaction between these factors ($F(1, 23) = 21.41$, $p < 0.001$, $\eta_p^2 = 0.48$). This reflects significantly slower responses for the semantic task in comparison to the episodic task for correct new responses ($p < 0.001$) and slower for correct when old than correct when new responses for the episodic task ($p < 0.001$).

Table 20. Reaction times (ms) for each task on switch, stay and stay+1 trials in Experiment 5. Standard deviations are in parentheses.

	Switch	Stay	Stay+1
Episodic task:			
Correct source	1542 (405)	1505 (383)	1447 (395)
Correct new	1318 (289)	1241 (259)	1180 (277)
Semantic task:			
Correct old	1537 (345)	1524 (410)	1451 (312)
Correct new	1529 (343)	1469 (375)	1454 (440)

7.5.2 Bayesian statistics

7.5.2.1 Behavioural analyses: replication of accuracy switch costs

A Bayesian replication test (Dienes, 2011; Verhagen & Wagenmakers, 2014) was carried out in order to investigate whether the behavioural results from Experiment 5 provide evidence for a replication of the accuracy switch costs in Experiment 1. The number of participants and the *t* values for the main effects of interest, in the original experiment (1) and replication experiment (5) are shown in **Table 21**.

Table 21. *t* values and sample sizes (*N*) for the accuracy switch costs in Experiment 1, and for the replication attempt (Experiment 5).

Study	*t* value	*N*
Experiment 1:		24
Old/new discrimination switch cost	-2.56	
Source accuracy switch cost	-2.77	
Experiment 5:		24
Old/new discrimination switch cost	-2.45	
Source accuracy switch cost	-0.51	

For Experiment 5 as a replication of the old/new discrimination switch cost in Experiment 1 the $BF_{r0} = 10.65$ (**Figure 24A**). This BF indicates that the data provide strong evidence for the alternative hypothesis. For a replication of the source accuracy switch cost the $BF_{r0} = 0.24$ (**Figure 24B**). This BF indicates that the data provide substantial evidence for the null relative to the alternative hypothesis. Further comment on the implications of these data is deferred until the outcomes of a second behavioural study are reported.
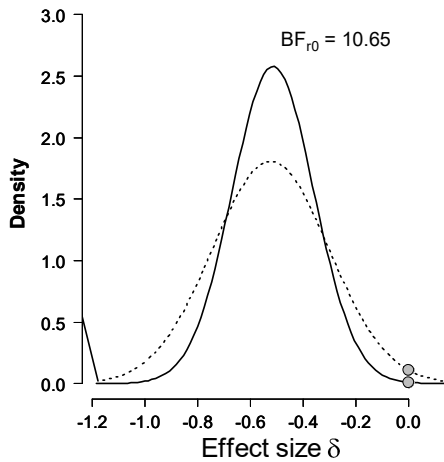
A. Old/new discrimination   **B.** Source
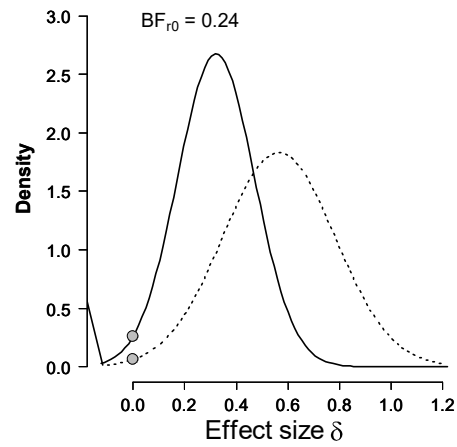
Figure 24. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the accuracy switch costs (A. Old/new discrimination; B. Source) in Experiment 1. In each panel, the dotted lines represent the posterior from the original study, which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Experiment 5) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

7.6   Experiment 6: Introduction

This experiment was conducted to investigate whether switch costs in memory tasks include an 'alternating cost' as a result of alternating between tasks. Alternating costs have been investigated in experiments where participants switched between completing two episodic memory tasks and also completed blocks of only one task-type (Herron & Wilding, 2006b; J. D. Johnson & Rugg, 2006; Werkle-Bergner et al., 2005).

J. D. Johnson and Rugg (2006) reported an accuracy cost for only some conditions in an alternating list compared to a blocked list. Werkle-Bergner et al. (2005) reported an old/new discrimination advantage in blocked lists, which was carried mainly by correct rejections. However, in Herron and Wilding (2006b) accuracy was equivalent across three experiments where two were alternating and one was a blocked task design.

Experiment 6 was designed to investigate whether additional alternating costs are evident when participants switch between completing an episodic memory task and a semantic memory task, which has not been reported to date. In keeping with some of the outcomes described (Herron & Wilding, 2006b; J. D. Johnson & Rugg, 2006; Werkle-Bergner et al., 2005), and in keeping with findings in the broader task-switching literature (Marí-Beffa, Cooper, & Houghton, 2011; Poljac, Koch, & Bekkering, 2009; Rubin & Meiran, 2005), it was anticipated that there would be alternating costs in this experiment.

7.7   Experiment 6: Method

7.7.1   Participants

A sample size of 24 participants was decided *a priori* based on power analyses (see **Appendix A**) and counterbalancing considerations. Data were collected from 25 participants, and the data from one participant were excluded due to an old/new discrimination score falling below 0.10. Thus, the data from 24 participants (M_{age} = 21, range = 18-23, 21 female) were included in the analyses. Participants were paid £6 per hour.

7.7.2   Design

The stimuli were the same 384 black line drawings of objects (Szekely et al., 2004) used in Experiment 5. The objects were presented following the same protocol as in Experiments 1-5.

As in Experiment 5 there were 128 objects in each semantic category, classified according to where they were usually found: inside, outside or both. For the semantic classification, the mean inter-rater reliability of three raters was 0.70. As in Experiment 5 the experiment comprised eight study-test cycles, and the 128 stimuli from each semantic category were randomly assigned to one of eight lists. Thus, each list contained 48 objects: 16 from each semantic category.

At study, 24 of the objects were either presented inside or outside an abstract outline of a building, and participants were required to indicate whether the object appeared inside or outside as in Experiments 1-5 (see **Figure 6**. Experiment 1. p. 66).

At test, the 24 objects from the preceding study phase were randomly intermixed with 24 unstudied objects. As in the previous experiments, each test object was preceded by one of two preparatory cues (X/O: presented in the centre of the screen), indicating which task participants were to prepare to complete. A capital 'X' directed participants to prepare for the episodic task, where they were to retrieve the prior study location of the object. A capital 'O' directed the participants to prepare for the semantic task. This task required the retrieval of information about the usual location of the object, regardless of the study phase, as in Experiment 1. However, in this experiment the cue-type was either presented for two consecutive trials (switch and stay trials: as in Experiment 1-3), or the cue-type remained the same throughout the whole test block. Participants were told before each block which task they would be completing. There were four study-test cycles for each block-type (blocked or alternating test phases). For blocked, participants completed the episodic task throughout two study-test cycles, and the semantic task throughout the other two. Whether the object appeared on a switch or stay trial, the block-type (blocked/alternating), block-type order (no more than two blocked or alternating phases in a row), and the old/new status of the object were counterbalanced across participants. Practice blocks were conducted for each block-type (and task) to familiarise participants with the experiment demands. As in Experiment 1, during the test phase responses were made using the same fingers as at study, with the addition of the index finger of the other hand to indicate 'new' or 'both', for the episodic and the semantic task respectively. The hands used for the judgments were counterbalanced across participants.

7.7.3   Procedure

The sequence within each trial was as in Experiments 1-3 and 5 (see **Figure 6**. Experiment 1. p. 66). Trials on which responses were faster than 300ms or slower than 4000ms were counted as errors and excluded from the behavioural analyses (0.9% of the trials).

7.8 Experiment 6: Results

7.8.1 Frequentist statistics

7.8.1.1 Behavioural analyses

During the study phases participants correctly responded 'inside' or 'outside' on 98% of trials. **Table 22** shows the response accuracy data, and **Table 23** shows the reaction time data, for the test phases (alternating (switch and stay) and blocked).

Table 22. Response accuracies for each task in the alternating (switch and stay trials) and blocked phases of Experiment 6. Standard deviations are in parentheses.

	Switch	Stay	Blocked
Episodic task:			
Old/new discrimination (*Pr*)	0.61 (0.25)	0.70 (0.18)	0.71 (0.23)
P(correct source)	0.82 (0.14)	0.80 (0.14)	0.83 (0.12)
Correct rejection	0.80 (0.14)	0.86 (0.11)	0.88 (0.12)
Semantic task:			
Correct classification	0.70 (0.10)	0.71 (0.12)	0.70 (0.08)

Table 23. Reaction times (ms) for each task in the alternating (switch and stay trials) and blocked phases of Experiment 6. Standard deviations are in parentheses.

	Switch	Stay	Blocked
Episodic task:			
Correct source	1462 (379)	1420 (339)	1259 (289)
Correct new	1265 (337)	1160 (340)	977 (298)
Semantic task:			
Correct old	1545 (357)	1405 (320)	1192 (244)
Correct new	1363 (296)	1344 (301)	1206 (226)

7.8.1.1.1 Alternating task phases

For response accuracy in the alternating task phases, discrimination scores (discrimination index: $Pr = p$(hit) - p(false alarm), Snodgrass & Corwin, 1988) for the episodic task were above zero for both trial-types (switch: 0.61, stay: 0.70) and higher on stay trials than on switch trials ($t(23) = 3.13$, $p < 0.05$, $d_z = 0.64$, 74% CL). The conditional probabilities of correct source judgments collapsed across the inside/outside dimension were

reliably above chance in both cases (switch: $t(23) = 11.22$, $p < 0.001$, $d_z = 2.29$, 98% CL; stay: $t(23) = 10.64$, $p < 0.001$, $d_z = 2.17$, 99% CL). Source accuracies in the episodic task and the semantic task accuracies were not statistically different across trial-types.

A 2x2x2 ANOVA was conducted on the reaction times for response accuracy categories (correct episodic source or semantic categorisation: when old/when new) separated according to task (episodic/semantic) and trial-type (switch/stay). Main effects of trial-type ($F(1, 23) = 18.44$, $p < 0.001$, $d_z = 0.88$, 81% CL), task ($F(1, 23) = 9.44$, $p < 0.05$, $d_z = 0.63$, 73% CL) and response accuracy ($F(1, 23) = 35.37$, $p < 0.001$, $d_z = 1.21$, 89% CL), were moderated by an interaction between all three factors ($F(1, 23) = 4.72$, $p < 0.05$, $\eta_p^2 = 0.17$) (there was also a task by response accuracy interaction: $F(1, 23) = 5.82$, $p < 0.05$, $\eta_p^2 = 0.20$).

Looking at each task separately there were main effects of trial-type ($F(1, 23) = 5.00$, $p < 0.05$, $d_z = 0.46$, 68% CL) and response accuracy ($F(1, 23) = 28.56$, $p < 0.001$, $d_z = 1.09$, 86% CL) for the episodic task: reflecting slower responses for switch in comparison to stay, and for correct when old than correct when new responses. For the semantic task there were also main effects of trial-type ($F(1, 23) = 10.47$, $p < 0.05$, $d_z = 0.66$, 75% CL) and response accuracy ($F(1, 23) = 16.92$, $p < 0.001$, $d_z = 0.84$, 80% CL); however, these were moderated by a trial-type by response accuracy interaction ($F(1, 23) = 4.51$, $p < 0.05$, $\eta_p^2 = 0.16$), reflecting significantly slower responses for switch than stay trials for correct when old ($p < 0.001$) but not correct when new responses.

7.8.1.1.2 Blocked task phases

For response accuracy in the blocked task phases, the discrimination score (discrimination index: $Pr = p$(hit) - p(false alarm), Snodgrass & Corwin, 1988) for the episodic task was 0.71. The conditional probability of a correct source judgment collapsed across the inside/outside dimension was reliably above chance (0.83, $t(23) = 12.86$, $p < 0.001$, $d_z = 2.63$, 99.6% CL). Semantic task accuracy was 0.70.

A 2x2 ANOVA was conducted on the reaction times for response accuracy categories (correct episodic source or semantic categorisation: when old/when new) separated according to task (episodic/semantic). A main effect of response accuracy category ($F(1, 23) = 20.88$, $p < 0.001$, $d_z = 0.93$, 82% CL) (and a trend for a main effect of task: $F(1, 23) = 3.92$, $p = 0.06$,

$d_z = 0.40$, 66% CL) was moderated by a task by response accuracy category interaction ($F(1,$ 23) = 29.78, $p < 0.001$, $\eta_p^2 = 0.56$), reflecting significantly slower responses for the semantic task in comparison to the episodic task for correct new responses ($p < 0.001$) and slower for correct when old than correct when new responses for the episodic task ($p < 0.001$).

7.8.1.1.3 *Blocked versus alternating task phases (alternating costs)*

Comparisons between stay trials and performance in the blocked phases indicated no differences according to the accuracy measures (old/new and source accuracy for the episodic task, and semantic task accuracy). However, for reaction time, a 2x2x2 ANOVA including the factors of trial-type (here: blocked/stay trials), task (episodic/semantic), and response accuracy (correct episodic source or semantic categorisation: when old/when new) indicated a main effect of trial-type ($F(1, 23) = 44.74$, $p < 0.001$, $d_z = 1.37$, 91% CL) reflecting faster responses during the blocked test phases in comparison to stay trials in the alternating task phases.

There was also a trend for a trial-type by task by accuracy interaction ($F(1, 23) =$ 3.72, $p = 0.066$, $\eta_p^2 = 0.14$), as well as main effects of task ($F(1, 23) = 6.34$, $p < 0.05$, $d_z =$ 0.51, 70% CL) and response accuracy ($F(1, 23) = 29.66$, $p < 0.001$, $d_z = 1.11$, 87% CL), which were moderated by a task by response accuracy interaction ($F(1, 23) = 28.16$, $p <$ 0.001, $\eta_p^2 = 0.55$). This reflects significantly slower responses for the semantic task in comparison to the episodic task for correct new responses ($p < 0.001$) and slower for correct when old than correct when new responses for the episodic task ($p < 0.001$).

7.8.2 Bayesian statistics

7.8.2.1 Behavioural analyses: replication of accuracy switch costs

A Bayesian replication test (Dienes, 2011; Verhagen & Wagenmakers, 2014) was carried out in order to investigate whether the behavioural results from Experiment 6 provide any evidence for a replication of the accuracy switch costs in Experiment 1. The number of participants and the *t* values for the main effects of interest, in the original experiment (1) and replication experiment (6) are shown in **Table 24**.

Table 24. *t* values and sample sizes (*N*) for the accuracy switch costs in Experiment 1, and for the replication attempt (Experiment 6).

Study	*t* value	*N*
Experiment 1:		24
Old/new discrimination switch cost	-2.56	
Source accuracy switch cost	-2.77	
Experiment 6:		24
Old/new discrimination switch cost	-3.13	
Source accuracy switch cost	0.59	

For Experiment 6 as a replication of the old/new discrimination switch cost in Experiment 1 the BF_{r0} = 42.48 (**Figure 25A**). This BF indicates that the data provide very strong evidence for the alternative hypothesis. For a replication of the source accuracy switch cost the BF_{r0} = 0.06 (**Figure 25B**). This BF indicates that the data provide strong evidence for the null relative to the alternative hypothesis.
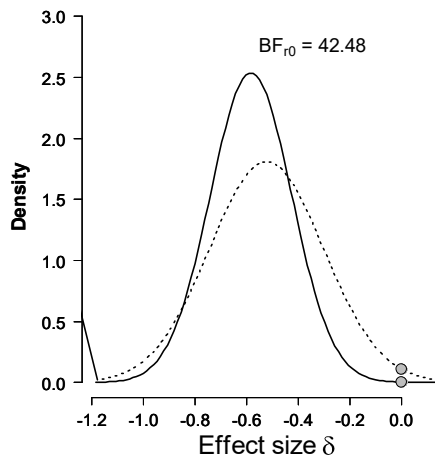
A. Old/new discrimination **B.** Source
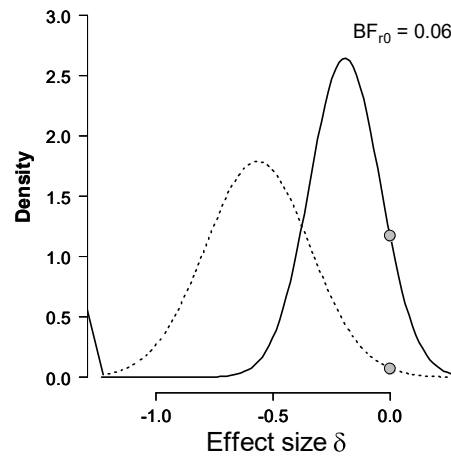
Figure 25. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the accuracy switch costs (A. Old/new discrimination; B. Source) in Experiment 1. In each panel, the dotted lines represent the posterior from the original study, which is used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Experiment 6) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

7.9   Experiments 5 & 6: Discussion

7.9.1   Experiment 5

Experiment 5 was conducted to investigate the time course of the behavioural costs identified previously. The design was the same as in Experiment 1, except that here three trials of the same task were completed in a row before participants were required to switch tasks. This design enabled investigation of whether there were any further improvements in performance on the third successive trial of the same task.

Commonly reported switch costs for reaction times on switch trials were replicated here. There were also costs for reaction times on stay trials, with faster responses on average for stay+1 trials. Old/new discrimination costs were, however, restricted to switch trials. In addition, there was a small but significant improvement in accuracy for the semantic task across switch to stay+1 trials.

Interestingly, the source accuracy switch cost for the episodic task, that was evident in Experiments 1 and 4, was not replicated here. In Experiments 1, 4, and 5, the content was matched across the tasks, as both tasks required the retrieval of 'location' information. It was suggested earlier that the source accuracy cost evident in Experiments 1 and 4 arose due to matching the content of retrieval across the tasks, and thus there was greater response overlap between the tasks than in Experiments 2 and 3. It was proposed that when there is more response set overlap (Experiments 1 and 4) there are increased processing demands that increase the likelihood of costs for source accuracy. However, it is possible that here an additional factor influenced the level of interference from response competition anticipated with matched content. In the current experiment, introducing an additional task repetition may have further influenced the predictability of the sequence. With a longer run of three trials of the same task perhaps participants became more aware of the task sequence, and, thus, were able to prepare for the task switch and compensate for the response set overlap that could potentially occur.

7.9.2 *Experiment 6*

Experiment 6 was conducted to investigate whether there were alternating costs above and beyond the switch costs identified in Experiments 1-5. The design was the same as in Experiment 1, except that in some blocks in this experiment participants completed the same task throughout the whole test block, and in others participants completed two trials of each task in succession. Performance measures were contrasted for the blocked and alternating task phases, as well as within the alternating phases (switch and stay trials).

In accordance with the previous literature switching costs for reaction time were observed (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). There was also a switch cost for old/new discrimination. In this experiment, as in Experiment 5, there were no switch costs for source accuracy despite the contents of retrieval being matched across the tasks. It is possible that task confusion was minimised in this experiment, as per the earlier argument, due to completing entire blocks of each task during the blocked task phases.

Moreover, when comparing the blocked task phases with stay trials from the alternating task phases there was no evidence for additional costs for the accuracy of the judgments. There were, however, additional costs for the reaction times of the judgments.

Participants were significantly faster during the blocked trials than stay trials in the alternating task phases, demonstrating additional alternating costs above and beyond switch costs for episodic memory retrieval.

7.9.3 Summary

In summary, Experiment 5 indicated that the accuracy costs identified previously are restricted to switch trials. Experiment 6 demonstrated that there are also costs associated with the general requirements to switch tasks. In addition, there were differences according to reaction time and response accuracy: there was a further significant improvement in reaction time for stay+1 trials (Experiment 5) and a cost over and above the switch costs was demonstrated for reaction times only (Experiment 6). The outcomes provide a starting point for understanding the boundary conditions for different kinds of switch costs in memory tasks. These will be considered in the **General discussion**.

8. CHAPTER 8. General discussion

8.1 Introduction

The work contained in this thesis was designed to contribute to the understanding of how healthy human individuals are able to selectively retrieve the information required to make accurate episodic memory judgments. Retrieval control operations were investigated under task-switching demands to explore the conditions necessary for the adoption of memory task-sets; the assumption being that one of the functions of sets is to facilitate retrieval processing (Herron & Wilding, 2004, 2006a). Electrophysiological and behavioural measures were collected as participants completed memory tasks in which cues were provided to signal the kinds of memory judgments that were to be made. Identifying ERP markers of distinct processing stages and investigating the behavioural benefits that are conferred when these indices are observed, is a means of informing our understanding of information processing models of memory.

Tulving (1983) proposed that in order to engage in episodic memory retrieval one must be in a particular cognitive state - retrieval mode. It was proposed that adopting this state allowed for inputs from the environment to be treated as cues for episodic memory retrieval (Tulving, 1983). Later Rugg and Wilding (2000) suggested that in order to demonstrate neural activity which could be associated with this task-set, there were three criteria that must be met. Firstly, the neural activity should be time-locked to the onset and maintained for the duration of an episodic retrieval task. Secondly, it should be evident when contrasting the activity observed during episodic retrieval tasks in comparison to non-episodic tasks. Thirdly, it should not vary across different episodic retrieval tasks.

As reviewed in **Chapter 1**, the studies in which haemodynamic imaging measures have been used to investigate retrieval mode fall short at meeting some or all of these criteria. Fundamentally, this is due to the lack of temporal precision offered by haemodynamic techniques. However, the temporal resolution of ERPs permits delineation of processing stages in real-time, and in the experiments reported here ERPs were employed in order to assess neural activity indexing preparatory processes that are engaged before retrieval cues (test items) are encountered.

In previous electrophysiological studies in which preparation for episodic memory retrieval was investigated, more positive-going activity was evident at right-frontal scalp sites following preparatory cues indicating an upcoming episodic task in comparison to a semantic 'baseline' task (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). In these experiments participants were cued on a trial-by-trial basis, and the divergence was obtained on stay trials only: the second successive trial of the same task.

This finding was interpreted as indicating that this ERP effect reflects the initial adoption and/or maintenance of an episodic retrieval set (that takes at least one trial to be engaged): that is, it is an index of retrieval mode (Herron & Wilding, 2004; Morcom & Rugg, 2002). The results in Herron and Wilding (2006a) allowed for development of this account, prompting claims that the ERP index reflects the initial adoption and/or configuration of mode rather than the maintenance of it. Herron and Wilding (2004) - in keeping with the criteria identified by Rugg and Wilding (2000) - showed that the effect is insensitive to episodic task demands.

The starting point for the work in this thesis was the observation of a potential confound in the studies described above. The manipulation employed most often to enable a separation of preparatory activity was the use of cues signalling participants should prepare to complete either an episodic or a semantic retrieval task. In these studies, however, the content of the information to be acted on also differed. For example, in Herron and Wilding (2006a), the episodic task required judgments about the location in which words had been shown at study, while the semantic task required participants recalling whether the object denoted by the word could move by its own accord (see also Herron & Wilding, 2004; Morcom & Rugg, 2002). Thus, it is possible that the differences between contents were contributors to the divergences seen across preparatory cue-types, which have commonly been assumed to reflect differences according to the kind of retrieval that is being prepared for.

8.2   *Summary of ERP results: Experiments 1-4*

In Experiment 1 of this thesis, where attempts were made to match the content of retrieval across two tasks, there was no evidence for the previously reported neural signature that has been linked to retrieval mode. This result is consistent with the view that the previous findings were due to content differences rather than preparation for distinct kinds of retrieval

from what are often assumed to be separable memory systems (Tulving, 1972). In order to assess this possibility, using the same episodic task that was employed in Experiment 1, in Experiment 2 the content of retrieval was unmatched across the tasks. There remained, however, no evidence for the previously reported neural signature. In Experiment 3 the stimuli were changed to words rather than pictures. The motivation for this manipulation was the assumption that it was possible that the index identified previously was a material-specific effect: the effect had been observed in studies only where verbal stimuli were employed (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). Again, there was no evidence for the previously reported neural signature linked with retrieval mode on stay trials. Experiment 4 was conducted, and in this experiment the task cue sequence was unpredictable and the RCI was shortened to 500ms. These manipulations were introduced following consideration of the task structure, guided by findings in task-switching designs in cognitive domains other than memory (Meiran et al., 2000; Monsell et al., 2003; Rogers & Monsell, 1995). Again, in Experiment 4 there was no evidence for the effect identified in several previous experiments on stay trials.

However, the preparatory activity associated with each cue-type did diverge in Experiments 1 and 4 on the switch trials. In Experiment 1 on switch trials there was more positive-going activity following the semantic than the episodic task cue at right-frontal scalp sites, and following the episodic than the semantic task cue at left-frontal scalp sites. In correspondence with Experiment 1, in Experiment 4 the preparatory activity also diverged on the switch trials. The divergence in Experiment 4, however, was in the direction obtained on stay trials in previous published studies: more positive-going for the episodic task at right-frontal scalp locations (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). Potential explanations for these differences will be discussed in the following subsections.

Firstly, implications of the ERP results from Experiments 1-4 will be discussed, followed by the implications from the outcomes of a recently published study (Evans, Williams, & Wilding, 2015). A summary and discussion of related accounts for the behavioural data from Experiments 1-6 will follow; then additional insights from ERP old/new and repetition effects, and a consideration of the use of Bayesian analyses in ERP studies. The final section includes consideration of the utility of the concept of retrieval mode, and possible mechanisms by which mode influences retrieval, future directions for this work, and, finally, concluding comments.

139

Perhaps the most interesting ERP finding from the work contained within this thesis is the discrepancy between the results obtained in Experiments 1-3 versus Experiment 4. Initially, it was not anticipated that ERP indices associated with preparation for episodic memory retrieval would be identified only under certain conditions according to the predictability of the cue sequence and the RCI that was employed (Herron & Wilding, 2004, 2006a, 2006b; Morcom & Rugg, 2002).

However, on consideration of the extant studies in which a putative index of retrieval mode was identified it was evident that the task cue sequence was unpredictable (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). The unpredictable nature of the sequence means that participants would not have been able to prepare for the next trial before the preparatory task cue. In Experiments 1-3, by contrast, the task cue sequence was predictable, so in principle preparation for the next trial could have commenced as soon as a response was made on the preceding trial. Moreover, in Experiments 1-3 the RCI was 1200ms. It is possible that an interval of this length enabled participants to initiate preparation for the forthcoming trial successfully in advance of the preparatory task cue. If this was the case, then neural activity time-locked to the cues might well fail to detect low frequency preparatory activity. This is a possible explanation for the absence of right-frontal activity linked to retrieval mode in Experiments 1-3.

Experiment 4 was conducted to investigate this possibility. The task cue sequence was unpredictable, and the RCI was shortened to 500ms. All other parameters remained the same as Experiment 1. In Experiment 4, the direction of the right-frontal divergence was the same as obtained in the published studies identifying an index on stay trials: more positive-going for the episodic than the non-episodic 'baseline' task. Thus, the lack of evidence for the episodic task right-frontal positivity in Experiments 1-3 is potentially due to the longer inter-trial interval and the predictability of the task cue sequence, enabling preparation before the preparatory cue, and thereby precluding the opportunity to observe the activity when it was time-locked to that cue.

Critically, however, this right-frontal divergence was observed on switch trials in Experiment 4. It is possible that the activity diverged on switch trials, rather than stay trials,

due to matching the content across the tasks. Perhaps this decreased the degree of cognitive reconfiguration required when switching between them. This explanation has been offered previously to explain why indices of retrieval orientation emerge on switch as opposed to stay trials (Herron & Wilding, 2006b; J. D. Johnson & Rugg, 2006). The authors suggested that when participants switch between two episodic tasks the degree of reconfiguration is smaller than when switching between an episodic and non-episodic task. When tasks share task-set components it is possible that resources can be mobilised more easily, and, therefore, the ERP index associated with the initial adoption of the task-set will be evident earlier.

An additional question is what the divergence obtained in Experiment 1 reflects, if participants were preparing in advance of the preparatory task cues in Experiments 1-3. It was initially considered that perhaps the effect represents suppression of the conflicting responses when the content was matched across the tasks (see *Experiment 2: Discussion*). However, the content was also matched across the tasks in Experiment 4. Perhaps, in Experiment 1, after the advanced preparation for the required task-set, suppression of the conflicting task-set or information followed.

In summary, in contrast to Experiments 1-3, in Experiment 4 there was more positive-going ERP activity following the episodic in comparison to the non-episodic task cues at right-frontal scalp sites. One explanation for this is that in Experiments 1-3 participants were preparing before the preparatory task cue as the sequence was predictable and they had time to, whereas in Experiment 4 the sequence was unpredictable and the inter-trial period was reduced: preventing preparation before the preparatory cue. Thus, factors not previously considered in regard to preparatory task-sets for episodic memory retrieval perhaps have more of a role to play, in combination, than initially expected. These findings highlight the need to consider various task design factors to constrain explanations for when preparation for episodic retrieval can be observed, and, consequently, to understand the benefits that preparation for episodic retrieval afford.

A different starting point, however, is to question the reliability of the previous findings, and to ask whether the consistent pattern of findings in Experiments 1-3 forms the basis for arguing that there is not in fact a stable and reliable index of retrieval mode. It is likely that, were the data in this thesis and the data reported previously to be submitted to a meta-analysis, the overall evidence for a robust and topographically consistent signature of

141

retrieval mode would be weak at best. A counter-argument, however, is that it is debatable whether combining studies solely on the basis of what they purport to measure, as opposed to on the basis of designs that enable that measurement, is an optimal way forward. In addition, it is also important to note that the outcomes of the Bayesian analyses provided evidence to suggest that in previously published studies there were not substantive concerns about power. Furthermore, across the set of studies in this thesis there was a marked degree of consistency in the behavioural outcomes, notably with respect to the reaction time switch costs. These data are also discussed elsewhere in this chapter (see section 8.5), primarily with respect to the question of how sets might influence retrieval, but for present purposes the behavioural outcomes are consistent with the view that participants did adopt a task-set, with the attendant costs reflecting processes linked to disengaging or initiating a new set. The behavioural data alone cannot distinguish between accounts that emphasise task-set inertia as opposed to reconfiguration (see **General introduction** section 1.3), but what is notable are the consistencies in the behavioural data for reaction times, both within this thesis and relative to other published studies. These of course contrast with the inconsistences in the electrophysiological data as already described. A parsimonious interpretation of these disparities is that the issue in Experiments 1-3 is one of task design, such that design elements precluded observation of a neural signature of a process that was in fact engaged to some extent. This argument receives support from the reaction time data and the ERP findings for the preparatory phase in Experiment 4, as well as the results of another study discussed in the section directly below.

8.4   *Implications of Evans, Williams, & Wilding (2015)*

The data in Experiment 4 broadly replicate those in a recently published experiment by Evans, Williams, et al. (2015). The similarities as well as the differences between the designs of these two experiments provide additional insights into indices of preparation for episodic retrieval. Evans, Williams, et al. (2015) looked at the ERP activity elicited when participants were cued trial-by-trial to complete an episodic versus a perceptual task. The task contents were highly similar, as they both involved location judgments. For the episodic task participants indicated the screen location of the word from the prior study phase (left/right/new). The perceptual task was the 'baseline' task in this instance (rather than the semantic tasks that are commonly used), and participants were required to indicate the current screen location of the word (top/middle/bottom).

In line with previous studies, there was more positive-going right-frontal activity when preparing for the episodic in comparison to the non-episodic baseline task. However, as in Experiment 4 of this thesis, the divergence in ERPs following the preparatory task cues was evident on switch trials only. Bayes Factors (BFs) were conducted for a replication of Evans, Williams, et al. (2015), and for Experiment 4 the BF was in favour of the alternative hypothesis (see *Appendix C*). Thus, Experiment 4 provides evidence for a replication of the ERP findings reported by Evans, Williams, et al. (2015): there was more positive-going right-frontal activity for the first (switch) trial of an episodic memory task following task cue presentation (800-1900ms). Critically, however, in Experiment 4 the evidence for a comparable signal on switch trials was obtained when the two tasks required either episodic or semantic retrieval, when the content was matched across the tasks.

Thus, there is no reason to believe that the effect identified by Evans, Williams, et al. (2015) was evident on switch trials because of the perceptual baseline task. In addition, Experiment 4 was the first demonstration of this divergence for pictorial stimuli. Thus, there is no reason to believe that the right-frontal positivity identified previously is a material-specific effect, or that there would be qualitative differences in preparatory neural activity associated with pictorial and verbal material as has been previously suggested (Wilckens et al., 2011).

Moreover, in both Evans, Williams, et al. (2015) and Experiment 4 the content was matched across episodic and non-episodic tasks (location information). The presence of the key divergence on switch trials can therefore be explained by the proposal that the degree of reconfiguration is smaller when switching between tasks involving similar or matched content (Evans, Williams, et al., 2015).

In Evans, Williams, et al. (2015) an alternative explanation for the presence of the modulation on switch trials was also suggested. It was proposed that the predictability of the task cue sequence could account for the emergence of the index on switch trials. In Evans, Williams, et al. (2015) the order of the task-switches was consistent: there were always two trials of each cue-type in succession. In addition, as there were only two tasks, the task that participants were switching to was also predictable (in line with EEG Experiments 1-3 of this thesis; however, here the RCI was 500ms - thus, perhaps participants did not have enough time to prepare during this short inter-trial period). As already mentioned, this is in contrast

143

to the other studies in the literature in which ERP indices of retrieval mode have been reported (Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). In the previous studies the requirement of a task switch, or the task that participants were switching to, was not predictable. However, in Experiment 4 of this thesis the task cue sequence was unpredictable, and more positive-going right-frontal activity was observed following the episodic task cue in comparison to the non-episodic (semantic) task cue on switch trials. Thus, in light of the results of Experiment 4, the predictability of the sequence as an explanation for the switch trial onset of the index in Evans, Williams, et al. (2015) is challenged.

8.5   *Behavioural results and implications: Experiments 1-6*

The behavioural task-switching results from the experiments contained in this thesis are relevant to the question of the benefits that are conferred by the successful adoption of a task-set. In Experiments 1-4 there were consistent reaction time switch costs, in line with the findings in published studies (Evans, Williams, et al., 2015; Herron & Wilding, 2004, 2006a; Morcom & Rugg, 2002). Across Experiments 1-4 there were also indications of switch costs for old/new discrimination in the episodic task. In addition, in Experiments 1 and 4, where the content was matched across the tasks, there were switch costs for the accuracy of the source judgments. In section 8.3 above the interpretations of the reaction time switch costs in these studies, and their relevance for constraining interpretations of the ERP preparatory data, were discussed. The behavioural outcomes also prompted the design of two further experiments.

In these, the boundary conditions for when behavioural switch costs occur were investigated. In Experiment 5 additional stay+1 trials were included to investigate whether the costs extended beyond switch trials. In Experiment 6 performance measures in 'blocked' and 'alternating' task phases were compared, to investigate whether the switch costs reflect in part an 'alternating cost' as a result of alternating frequently between tasks. In Experiment 5 the old/new discrimination cost was restricted to switch trials, however, there was further significant improvement in reaction times from stay to stay+1 trials. In Experiment 6 the key finding was that there was an additional cost associated with the general requirements to switch tasks, but this was observed for reaction times only. These results suggest that reaction time and response accuracy measures are not indexing precisely the same cognitive operations, although this depends upon an unlikely assumption that there is comparable

sensitivity in the reaction time and accuracy measures. In addition, in Experiments 5 and 6 there were no switch costs for source accuracy. Source accuracy switch costs are discussed first below followed by a consideration of old/new discrimination costs.

A preliminary explanation offered in **Chapter 4** (***Experiment 2: Discussion***) was that matched content across the tasks was accountable for the source accuracy cost in Experiment 1 and the absence of this cost in Experiment 2, where content was not matched. It was proposed that in Experiment 1 there may have been greater interference and additional demands resulting from response competition across the episodic and semantic tasks because the content was highly similar. This explanation also holds when considering Experiments 3 and 4. Content was only matched in the latter, and only in this experiment was a source accuracy cost observed.

However, the content was also matched in Experiments 5 and 6, and switch costs for source accuracy were not observed. It is possible that this was due to a combination of the increase in run length (the inclusion of stay+1 trials in Experiment 5), or completing entire blocks of each task (during the blocked task phases in Experiment 6) and the predictability of the sequences. These changes could allow participants to overcome any response conflict/task confusion somewhat more easily than was the case when the run length was shorter, or when the sequence was unpredictable. In addition, findings from the wider task-switching literature have indicated that with unpredictable sequences there is a more gradual approach to asymptotic performance than with predictable sequences (Monsell et al., 2003). This might also explain the large switch cost in source accuracy obtained in Experiment 4 when the task cue sequence was unpredictable.

To summarise, the behavioural data from these experiments indicate that how accurate people's source memory judgments are depends, in part, on what other tasks they are completing and the potential for response conflict/task confusion. Reaction time costs extend beyond the first trial of a task, and are additionally associated with the general requirements to alternate between tasks. However, people are both slower and less accurate when just starting an episodic memory task than when having completed the task for one trial.

These switch costs fit with earlier proposals that it takes at least one trial of a task before completely adopting a task-set (Düzel et al., 1999; Monsell, 2003; Morcom & Rugg,

145

2002). What is new about the findings in this thesis is the switch cost for old/new discrimination accuracy. This has not been reported in task-switching episodic memory studies previously (Evans, Williams, et al., 2015; Herron & Wilding, 2004, 2006a, 2006b; Morcom & Rugg, 2002; Wilckens et al., 2011). To explain the absence of an old/new discrimination cost, Morcom and Rugg (2002) suggested that the availability of the familiarity process was not affected by switching manipulations, and that recognition judgments can be based on familiarity. One explanation for the different findings in the experiments in this thesis is that source judgments were required, while Morcom and Rugg (2002) required only old/new judgments. It is possible that participants relied to a greater extent on recollection for old/new judgments because of this requirement, and recollection was less available on switch than on stay trials (for ERP data consistent with this account see Evans et al., 2012; and Wilckens et al., 2011). This account does not explain, however, the absence of old/new discrimination costs in other published studies in which source judgments as well as task-switches were required (Evans, Williams, et al., 2015; Herron & Wilding, 2006a).

It is also possible that the old/new switch costs can be accounted for by changes in decision-related processes (e.g. defaulting to responding 'old' when unsure), rather than the amount of mnemonic evidence associated with different items. In line with this, across some of the experiments there is evidence for an increased false alarm rate on switch trials relative to stay trials (see *Appendix D*: correct rejections). It is, however, difficult to investigate changes in decision-related processes using single-point measures (Dougal & Rotello, 2007; Rotello, Masson, & Verde, 2008). This is because discrimination scores may be inaccurate when there are differences in response bias across conditions.

One way to assess this possibility is to introduce confidence levels for each response on switch and stay trials. Introducing a subsequent confidence judgment into the task-switching design would allow one to assess whether the increased false alarm rate is attributable to defaulting old when unsure. This consideration is important because previously performance measure outcomes have been attributed to the successful adoption of a task-set, which when utilised means that one is more likely to recover task relevant information. However, if the switch cost is indeed driven more by the increased false alarm rate to new items, then, as stated, the reason for this could be operations that influence control processes

related to bias, rather than the recovery of information *per se*. Thus, further investigation into what is driving the switch costs observed would be beneficial.

There are also similarities between the findings on switch and stay trials here and the 'revelation effect' where participants are more likely to default to respond 'old' when there is an interpolated task between successive old/new judgments (Watkins & Peynircioglu, 1990). In the revelation effect paradigm participants either complete a series of successive trials where old/new judgments are required, or complete a different task before each judgment. Sometimes the conditions are blocked, and sometimes they are interspersed. Effectively these can be regarded as task-switching paradigms of a sort, and as a result the ways in which revelation effects have been explained is worthy of consideration. It has been broadly assumed that revelation effects emerge either because of changes in the familiarity of test items or changes in response bias (Hicks & Marsh, 1998).

In an ERP study designed to investigate explanations for this effect the ERP index of familiarity was of lower amplitude (for old as well as for new words) for trials where an anagram task was completed just before the test stimulus (Azimian-Faridani & Wilding, 2004). It was concluded that the reduction in familiarity for critical test items contributed to the revelation effect (Azimian-Faridani & Wilding, 2004). There are, however, data points from experiments in which Receiver Operating Characteristic (ROC) curves were plotted that challenge this proposal (Dougal & Rotello, 2007; Verde & Rotello, 2003, 2004).

Thus, future investigations of the reasons for the effects in Experiments 1-6 could include (i) confidence judgments, or (ii) exploration of the ERP correlates of retrieval processes. Designs of this kind are methodologically challenging because of the fact that (i) task-switching designs are already demanding for participants and an additional confidence judgment for each trial may prove difficult, and (ii) there are issues around balancing task duration and participant motivation to obtain sufficient trial numbers to investigate ERP indices of memory retrieval in task-switching designs. Some of these caveats apply also to the analysis of ERP old/new effects and repetition effects in the experiments in this thesis, with which the next section is concerned.

8.6   ERP old/new and repetition effects

Additional insights into how successful preparation influences retrieval can be gained by analysing ERP old/new effects elicited by the retrieval cues. A general assumption is that preparatory operations determine the subsequent processes that operate when a retrieval cue is encountered (Herron & Wilding, 2006b; Rugg & Wilding, 2000; Wheeler et al., 1997). Thus, the consequences of successfully adopting a task appropriate set may be revealed by the activity elicited in response to the test items (Herron & Wilding, 2006b). Examination of the ways in which the old/new effects vary for switch and stay trials offers a means to determine the influence of the switching demand on the successful adoption of a retrieval set, and the benefit successful adoption has on retrieval success. This approach has been utilised previously by Wilckens et al. (2011), Evans et al. (2012), and Evans, Herron, et al. (2015).

Wilckens et al. (2011) reported that the left-parietal ERP old/new effect was larger on episodic repeat (stay and stay+1 trials collapsed) than switch trials. Similarly, Evans et al. (2012) reported larger left-parietal old/new effects on stay+1 than stay and switch trials. These findings are consistent with the view that when the time available to adopt a set is increased the quality or volume of recovered episodic information also increases (Evans et al., 2012; Wilckens et al., 2011).

Evans, Herron, et al. (2015) demonstrated that ERP repetition effects were larger on switch than on stay trials in the task requiring perceptual judgments. Thus, task-irrelevant information was more available on switch than on stay trials. The results provide support for the 'task-set inertia' account of task-switching (Allport et al., 1994; Düzel et al., 1999; Wylie & Allport, 2000), although other processes may still contribute. In an additional analysis, Evans, Herron, et al. (2015) demonstrated that the magnitude of the repetition effects observed on the perceptual task trials was related to the reaction time switch cost: as the task-irrelevant activity reduced so did the reaction times. Thus, task-set inertia (as indexed by recovery of task-irrelevant information) was presumed to have a functional role in the behavioural perceptual task switch cost.

In the experiments contained in this thesis, analysis of the test item data was limited due to insufficient trial numbers for comparison across conditions of interest for a number of participants. However, with a subset of the participants, analyses of the processes related to

retrieval success (old/new and repetition effects) were conducted and are reported in **Appendix B**. The statistical outcomes for the test item ERP data from Experiments 1-4, however, do not offer much insight, in all likelihood due to the somewhat low participant and average trial numbers contributing to the conditions of interest. Overall the statistical evidence for changes in the magnitudes of effects was limited and variable across the experiments (see **Appendix B**). In Experiment 1 there was an indication that the repetition effect was larger on stay than on switch trials, while in Experiment 2 there was some evidence that the old/new effect was larger on switch than on stay trials. These partial outcomes make it difficult to use these data to draw strong conclusions about the ways in which preparation for retrieval influences subsequent retrieval processing.

8.7   *Use of Bayesian analyses*

Bayesian analysis is an alternative statistical approach which has not been adopted particularly widely; however, its use is becoming more prominent. This is due to concerns about the use of traditional NHST, and what are seen as flaws around using the arbitrary $p < 0.05$ cut-off, as well as the 'crisis of confidence' in experimental psychology and the neurosciences regarding replication failures perhaps due to underpowered initial studies and false-positive reporting due to the file drawer problem (Button et al., 2013; Dienes, 2011; Ioannidis, 2005; Rosenthal, 1979; Simmons et al., 2011). Throughout this thesis, Bayes factors were calculated in order to take up the opportunity to assess evidence in favour of the null or alternative hypothesis, and to determine whether the sample sizes employed were appropriate for the effect sizes of interest. For the replication test used here, the only input parameters required are the numbers of participants and the t values for the effect of interest, from both the original and replication experiments (Verhagen & Wagenmakers, 2014). Thus, the subjectivity in selection of the prior, and the researcher degrees of freedom, is reduced to prevent biased results.

The use of the Bayesian and power analyses in this thesis proved to be an effective means of guiding data acquisition via the stopping rule. This is a more principled approach than what is typically done, which is to base sample size on precedents in the existing literature. That is not to say that prior results should necessarily be challenged or considered circumspect, and it is likely that effects that have survived and been replicated over time are partly a consequence of a convergence on a set of broadly appropriate parameters. None the

less, having external validation is important, and one important outcome of the work here is an indication that (broadly) participant numbers were appropriate in previous published studies, and for the effects of interest in Experiments 1-6. This is an important outcome in and of itself, but it also emphasises that there may be a particular role for applying a stopping rule in initial replications of effects that are not well-established.

Additional new information provided by this approach is the indication of the strength of the evidence in Experiment 4 as a replication of Evans, Williams, et al. (2015). Experiment 4 only provided anecdotal evidence for the same effect having been identified on switch trials. This is information that would not have been available, at least in a numerical sense, via traditional approaches. The number of participants was increased from 24 to 32 in line with the Bayesian Stopping Rule, however, the results remained the same and data collection was terminated at this point.

One explanation for this outcome is that the evidence was not substantial because the average activity across the three right-frontal sites was compared for each task (as this was how the t value was obtained from the analysis conducted in the original experiment), and in Experiment 4 the divergence appears more focal and primarily for the most inferior right-anterior electrode site (F8). The exact same sites were submitted to analysis in accordance with the original experiment, however, this approach does not account for the variability evident in the literature. For instance, in the electrophysiological literature consideration is required for whether an effect is the same if it is largest at an adjacent electrode in a replication sample. For example, in Düzel et al. (1999) more positive-going activity was largest for the episodic task at the right fronto-polar electrode (Fp2). In Morcom and Rugg (2002), the effect had a diffuse, right central maximum (Fz, Cz, F4), in Herron and Wilding (2004) differences were greatest at the right-frontal mid-lateral electrode (F6), and in Herron and Wilding (2006a) at the right-frontal inferior electrode (F8). Given the susceptibility of ERP scalp distributions to variations in head shape and size and their consequences for specific electrode locations when electrode caps are employed, some variability in distributions is to be anticipated, but how to assess this directly remains unclear. The challenge extends to Bayesian analyses, as outlined below.

A specific difficulty is that it is not clear how precisely matched the Bayesian comparisons need to be, for instance when not entirely the same electrode montage was used, or whether one can restrict selection of sites to account for typical variability in exactly where on the scalp the effect was obtained. Restrictions on the use of Bayesian intuition, which would be necessary to prevent researcher bias, are not entirely clear. Here a stringent approach was adopted where the *t* values were obtained in exactly the same way as in the original study of comparison. It would have been possible, for example, to adopt a different metric, such as the site or sites with the largest divergence within a quadrant. This is of course partly dependent upon what information can be extracted from prior studies, and if nothing else what the preliminary analyses in this thesis have served to highlight are the range of questions that need to be considered if this approach is to be employed successfully.

8.8   *Utility of the concept of retrieval mode*

It is clear that investigating ERP indices of task-sets using task-switching designs has limitations. It has proved more difficult than anticipated to use these designs as a tool to inform models of memory. This is because the work in this thesis has demonstrated that there are other factors which were not originally considered to be of particular importance when designing studies that will develop an understanding of memory retrieval processing.

In this thesis task-switching designs were used to enable separation of the processes occurring before and after the presentation of the retrieval cue. The outcomes indicate that ERP indices are evident only under a certain set of conditions. From a pragmatic perspective the findings in Experiments 1-4 delineate the kinds of designs it is necessary to use to observe signatures of preparatory retrieval processes. This is important because observing effects of interest is a precursor to being able to use the effects to understand how preparation influences subsequent processing operations.

The explanation offered for the null results in Experiments 2 and 3 carries the assumption that preparation did in fact occur, but the trial designs in those experiments precluded observation of the effects of interest in the electrical record. If this account is correct, then the outcomes in these experiments do not speak to the question of whether there are in fact distinct signatures of retrieval mode alongside patterns of preparatory activity that vary according to specific retrieval demands (retrieval orientations). Moreover, these

151

outcomes limit the extent to which the electrophysiological data can add to what is known at the conceptual level about retrieval sets and their sequel. This is of course true only for the ERP data, and the new knowledge generated from the accuracy, most notably regarding old/new discrimination, is covered in section 8.5.

There is also, however, some new knowledge about the properties of retrieval sets that the ERP data offer. The outcomes in Experiments 1-4 are consistent with the view that there is an ERP signature of retrieval mode that has a right-frontal scalp distribution and an extended time course. The new information provided by the findings here concerns the period over which retrieval mode can be engaged. The results of Experiment 4, where the effect was evident on switch trials, are consistent with the view that the similarities between the contents of the tasks that people are switching to and from determine how quickly a relevant set can be adopted (Evans, Williams, et al., 2015). This has theoretical implications, because it can be interpreted as evidence that task-set configuration can be enabled in this way, hence underpinning the claim that reconfiguration is itself a contributor to the costs seen in retrieval tasks, most likely alongside a contribution from task-set inertia (Evans, Herron, et al., 2015).

A question that follows on from the identification of indices of preparatory retrieval processes is how they operate to benefit subsequent retrieval processing operations. The ERP findings reported in this thesis can be considered within the taxonomy of models (inhibitory and non-inhibitory) originally presented by Anderson and Bjork (1994) who introduced the notions of 'target-bias' and 'cue-bias'. 'Target-bias' models are inhibitory models of retrieval, whereby an activation-reducing inhibitory mechanism operates. In these models there are decrements in activation for the target item representation. 'Cue-bias' models are non-inhibitory models; where there are operations applied to a cue which effectively bias the way memory is searched. Thus, target-bias includes the mechanisms that act on memory representations directly to modulate the accessibility of them (Mecklinger, 2010). Cue-bias relates to the external retrieval cue, and the mechanisms that augment its processing (Mecklinger, 2010).

Cue- and target-biases may act as mechanisms by which retrieval mode influences subsequent retrieval. Thus, the neural signature of retrieval mode may index processes responsible for ensuring that cues will be subject to a certain kind of processing (cue-bias), and/or may index processes responsible for ensuring some representations are more

accessible than others (target-bias). Mecklinger (2010) put forward the idea that target-bias mechanisms are anticipatory and operate in response to the preparatory task cue in alternating retrieval demands, and that cue-bias mechanisms operate in response to the subsequently presented retrieval cue. Mecklinger (2010) noted that the strategy of target-bias mechanisms following preparatory task cues could be particularly efficient in situations where people are required to alternate between tasks with unpredictable retrieval demands. In addition, J. D. Johnson and Rugg (2006) stated that the effects elicited by preparatory task cues on switch trials reflect the differential processing associated with the target-bias strategy (as this is when the target-bias must be switched).

However, differences between ERPs elicited by preparatory cues, or following the retrieval cue (for correct rejections), do not allow one to separate cue- and target-bias. This is because differences between ERPs to new items may ensue following preparatory processes which operate directly on representations (target-bias), or, preparatory activity may index preparation to treat cues differently (cue-bias), and hence differences between ERPs elicited by new items. There is scope for investigation of these issues with fMRI, in so far as it is possible to identify regions responsible for supporting different kinds of memory and observe how their activity changes during preparation for retrieval of different kinds. There remain, however, the practical challenges of isolating activity related to preparation using fMRI.

8.9   *Future directions*

The work in this thesis has been successful at delineating the circumstances under which indices of retrieval mode can be observed, and providing indications that the time course over which preparation can occur depends upon the relationship between the contents of the episodic and non-episodic tasks that are employed.

In light of preparation time and the predictability of the task cue sequence proving important in designs of this kind, experimental designs in which cue sequences are unpredictable should prove fruitful for investigating preparatory processes. It may also prove useful to restrict the duration for participant responses so that preparation can be time-locked more directly to the task cue. Previously, when a 4000ms cue-item interval was employed it was decreased to 1500ms on a small number of trials (the catch trials in this instance) to encourage participants to prepare for the next trial as soon as the task cue was presented

(Herron & Wilding, 2006a). In addition, Herron and Wilding (2004) suggested that the lack of a response-deadline was accountable for the lack of switch costs to accuracy in their experiment: with a response-deadline the performance levels on switch trials may decrease, and the levels on stay trials may remain the same. A response-deadline could be utilised in future studies in order to ensure participants are motivated to prepare for the task directly after the task cue, so that preparation can be time-locked to it.

In terms of extensions to the work described here, the experiments in this thesis have thrown up several issues that might be addressed in new experiments. In Experiment 5 it was shown that the old/new discrimination cost was restricted to switch trials. In Experiment 5 the task cue sequence was predictable, thus, a further question of interest is whether an accuracy improvement is evident on the third trial of the same task when the task sequence is unpredictable. This could be achieved by a future experimental design with an unpredictable sequence and a sufficient number of stay+1 trials to allow investigation of this. As already mentioned, the broader task-switching literature has indicated that there is a more gradual approach to asymptotic performance with unpredictable sequences than with predictable demands (Monsell et al., 2003). Thus, perhaps with an unpredictable sequence there would be a cost for the stay trials also.

In addition, the concept of retrieval orientation is one to which considerable attention has also been paid, and one immediate question of interest is whether the same kinds of design restrictions apply to the circumstances under which indices of orientation will be identified. If precisely the same sets of parameters necessary for indexing retrieval mode are not necessary for indexing retrieval orientations, there is the opportunity to characterise how these two related classes of preparatory retrieval process differ.

Questions that are relevant to both mode and orientation include the specific benefits that preparatory retrieval sets confer. There is currently an incomplete picture of how successful adoption of a retrieval set influences subsequent retrieval processing. As already noted, there is some evidence that the quality or amount of recovered information available is diminished on switch trials, but this has not been observed consistently. This may reflect signal-to-noise issues, and/or some systematic differences between the conditions under which the quality or volume of information that is recovered can be influenced by preparation time and specific switching demands. Given the challenges in using ERPs to assess this, it

may be that the use of response-deadline procedures and confidence judgments will provide complementary insights. These manipulations both offer ways of understanding how recollection and familiarity might be affected by switching demands, and a related gap in the literature is the absence of studies in which task-switching manipulations have been employed alongside the remember/know procedure (Tulving, 1985; Yonelinas & Jacoby, 1995). The remember/know measure may reveal that switching manipulations influence recollection and familiarity differently. Yonelinas (2002) has noted the value of employing multiple measures to assess process engagement, and it is likely that this approach will be necessary to understand how switching demands influence the accuracy and time-course of memory judgments.

Remaining with functions of preparatory sets, it is broadly assumed that adopting mode can influence memory judgments positively. If this is the case, then the magnitude of an index might vary with measures of performance. While presenting design challenges, the ability to separate preparatory activity according to the accuracy and/or speed of subsequent judgments offers another means of investigating the benefits mode confers. This kind of experiment is of course based on the use of a design that elicits a robust index of retrieval mode, and the findings in this thesis go some way towards identifying the key design elements that will achieve that. Recognising these differences at the individual level, it may be that a fruitful approach is to assess the magnitudes of indices of preparation for retrieval under different levels of episodic task difficulty. It may be that when retrieval is easy - as indexed by good performance - there is little need to engage in advance preparation.

There is also scope to investigate the relationship between preparation for encoding and for retrieval. Otten and colleagues have reported their findings in studies where neural activity was recorded while participants prepared to encode different kinds of contents (Galli, Gebert, & Otten, 2013; Galli, Griffiths, & Otten, 2014; Galli, Wolpe, & Otten, 2011; Otten, Quayle, Akram, Ditewig, & Rugg, 2006; Otten, Quayle, & Puvaneswaran, 2010). Activity preceding a to-be-encoded item predicted whether it would be subject to a correct memory judgment on a subsequent test (Otten et al., 2006). It is possible that effective preparation for encoding and retrieval interact such that the former diminishes the need for the latter. Alternatively, effective preparation may be an individual trait, and if this is the case then markers of preparation at encoding and retrieval may be more likely to co-occur.

155

8.10 Concluding comments

Despite previous studies consistently demonstrating more positive-going neural activity during preparation for episodic memory retrieval, the experiments in this thesis demonstrate that the signature is not stable in time and across demands. The results indicate that the observation of the index is dependent on trial timing and task-sequence characteristics, and suggest that the trial onset of the signature depends on the degree of cognitive reconfiguration required. The index was identified on switch trials rather than stay trials (proposed as due to matching the content across the episodic and semantic tasks), when the inter-trial interval was short and the task cue sequence was unpredictable (preventing any advanced preparation). Thus, to extract a signal of retrieval mode to inform our understanding of the retrieval of information from episodic memory, future experimental design must take into account these findings. In addition, behavioural accuracy switch costs were evident which were not demonstrated previously, and their boundary conditions were identified (restricted to switch trials). These findings are relevant because memory retrieval is something we do interspersed, and often in parallel, with other tasks. Moreover, identifying ERP markers of distinct processing stages, and developing accurate information processing models of how memory control operates, is a precursor to understanding how episodic memory is affected in old age, as well as in disease and following selective brain damage (Wilding & Herron, 2006).

9. References

Abdi, H. (2010). The greenhouse-geisser correction. *Encyclopedia of research design.* Thousand Oaks, CA: SAGE Publications.

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: conscious and nonconscious information processing* (pp. 421-452). Cambridge, MA: The MIT Press.

Altmann, E. M. (2007). Comparing switch costs: alternating runs and explicit cuing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 475-483. doi:10.1037/0278-7393.33.3.475

Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long-term memory: a new taxonomy. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 265-325). San Diego, CA: Academic Press.

Azimian-Faridani, N., & Wilding, E. L. (2004). An event-related potential study of the revelation effect. *Psychonomic Bulletin & Review, 11*(5), 926-931. doi:10.3758/BF03196723

Bäckman, L., Almkvist, O., Andersson, J., Nordberg, A., Winblad, B., Reineck, R., & Långström, B. (1997). Brain activation in young and older adults during implicit and explicit retrieval. *Journal of Cognitive Neuroscience, 9*(3), 378-391. doi:10.1162/jocn.1997.9.3.378

Bayarri, M. J., & Mayoral, A. M. (2002). Bayesian design of "successful" replications. *The American Statistician, 56*(3), 207-214. doi:10.1198/000313002155

Bland, J. M., & Altman, D. G. (1994). Regression towards the mean. *BMJ: British Medical Journal, 308*(6942), 1499. doi:10.1136/bmj.308.6942.1499

Box, G. E. P. (1954a). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics, 25*(2), 290-302. doi:10.1214/aoms/1177728786

Box, G. E. P. (1954b). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *The Annals of Mathematical Statistics, 25*(3), 484-498. doi:10.1214/aoms/1177728717

157

Brain Products Brain Vision (UK) Ltd. (2014). *User manual for the 40 and 72-channel physiological measurement system QuickAmp-40 and QuickAmp-72* (pp. 3-16).

Brass, M., & von Cramon, D. Y. (2004). Decomposing components of task preparation with functional magnetic resonance imaging. *Journal of Cognitive Neuroscience, 16*(4), 609-620. doi:10.1162/089892904323057335

Bridger, E. K., Herron, J. E., Elward, R. L., & Wilding, E. L. (2009). Neural correlates of individual differences in strategic retrieval processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1175-1186. doi:10.1037/a0016375

Buckner, R. L. (1996). Beyond HERA: contributions of specific prefrontal brain areas to long-term memory retrieval. *Psychonomic Bulletin & Review, 3*(2), 149-158. doi:10.3758/BF03212413

Buckner, R. L. (2003). Functional-anatomic correlates of control processes in memory. *The Journal of Neuroscience, 23*(10), 3999-4004. doi:0270-6474/03/233999-06$15.00/0

Burgess, P. W., & Shallice, T. (1996). Confabulation and the control of recollection. *Memory, 4*(4), 359-411. doi:10.1080/096582196388906

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365-376. doi:10.1038/nrn3475

Cabeza, R., Kapur, S., Craik, F. I. M., McIntosh, A. R., Houle, S., & Tulving, E. (1997). Functional neuroanatomy of recall and recognition: a PET study of episodic memory. *Journal of Cognitive Neuroscience, 9*(2), 254-265. doi:10.1162/jocn.1997.9.2.254

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: an empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12*(1), 1-47. doi:10.1162/08989290051137585

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience, 2*(7), 671-676. doi:10.1038/10230

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.

Coles, M. G. H., Smid, H. G. O. M., Scheffers, M. K., & Otten, L. J. (1995). Mental chronometry and the study of human information processing. In M. D. Rugg & M. G.

H. Coles (Eds.), *Electrophysiology of mind: event-related brain potentials and cognition* (pp. 86-131). New York, NY: Oxford University Press.

Craik, F. I. M., & Lockhart, R. 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

Cumming, G. (2013). The new statistics: why and how. *Psychological Science, 25*(1), 7-29. doi:10.1177/0956797613504966

Dienes, Z. (2011). Bayesian versus orthodox statistics: which side are you on? *Perspectives on Psychological Science, 6*(3), 274-290. doi:10.1177/1745691611406920

Donaldson, D. I., Allan, K., & Wilding, E. L. (2002). Fractionating episodic memory retrieval using event-related potentials. In A. Parker, E. L. Wilding, & T. J. Bussey (Eds.), *The cognitive neuroscience of memory: encoding and retrieval* (Vol. 1, pp. 39-58). New York, NY: Psychology Press.

Donaldson, D. I., Petersen, S. E., Ollinger, J. M., & Buckner, R. L. (2001). Dissociating state and item components of recognition memory using fMRI. *Neuroimage, 13*(1), 129-142. doi:10.1006/nimg.2000.0664

Donders, F. C. (1868/1969). On the speed of mental processes. *Acta Psychologica, 30*(1), 412-431. doi:10.1016/0001-6918(69)90065-1

Dougal, S., & Rotello, C. M. (2007). "Remembering" emotional words is based on response bias, not recollection. *Psychonomic Bulletin & Review, 14*(3), 423-429. doi:10.3758/BF03194083

Dumermuth, G., Ferber, G., Herrmann, W. M., Hinrichs, H., & Künkel, H. (1987). Recommendations for standardization of data acquisition and signal analysis in pharmaco-electroencephalography. *Neuropsychobiology, 17*(4), 213-218. doi:10.1159/000118367

Düzel, E., Cabeza, R., Picton, T. W., Yonelinas, A. P., Scheich, H., Heinze, H. J., & Tulving, E. (1999). Task-related and item-related brain processes of memory retrieval. *Proceedings of the National Academy of Sciences, 96*(4), 1794-1799. doi:10.1073/pnas.96.4.1794

Düzel, E., Picton, T. W., Cabeza, R., Yonelinas, A. P., Scheich, H., Heinze, H. J., & Tulving, E. (2001). Comparative electrophysiological and hemodynamic measures of neural activation during memory-retrieval. *Human Brain Mapping, 13*(2), 104-123. doi:10.1002/hbm.1028

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*(3), 193-242. doi:10.1037/h0044139

Evans, L. H., Herron, J. E., & Wilding, E. L. (2012). Electrophysiological insights into control over recollection. *Cognitive Neuroscience, 3*(3-4), 168-173. doi:10.1080/17588928.2012.662217

Evans, L. H., Herron, J. E., & Wilding, E. L. (2015). Direct real-time neural evidence for task-set inertia. *Psychological Science, 26*(3), 284-290. doi:10.1177/0956797614561799

Evans, L. H., Williams, A. N., & Wilding, E. L. (2015). Electrophysiological evidence for retrieval mode immediately after a task switch. *Neuroimage, 108*(1), 435-440. doi:10.1016/j.neuroimage.2014.12.068

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi:10.3758/BF03193146

Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The long way from α-error control to validity proper problems with a short-sighted false-positive debate. *Perspectives on Psychological Science, 7*(6), 661-669. doi:10.1177/1745691612462587

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition, 28*(1-2), 3-71. doi:10.1016/0010-0277(88)90031-5

Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: a selective review. *Microscopy Research and Technique, 51*(1), 6-28. doi:10.1002/1097-0029(20001001)51:1<6::AID-JEMT2>3.3.CO;2-I

Friston, K. J. (2005). Models of brain function in neuroimaging. *Annual Review of Psychology, 56*(1), 57-87. doi:10.1146/annurev.psych.56.091103.070311

Galli, G., Gebert, A. D., & Otten, L. J. (2013). Available processing resources influence encoding-related brain activity before an event. *Cortex, 49*(8), 2239-2248. doi:10.1016/j.cortex.2012.10.011

Galli, G., Griffiths, V. A., & Otten, L. J. (2014). Emotion regulation modulates anticipatory brain activity that predicts emotional memory encoding in women. *Social Cognitive and Affective Neuroscience, 9*(3), 378-384. doi:10.1093/scan/nss145

Galli, G., Wolpe, N., & Otten, L. J. (2011). Sex differences in the use of anticipatory brain activity to encode emotional events. *The Journal of Neuroscience, 31*(34), 12364-12370. doi:10.1523/JNEUROSCI.1619-11.2011

Grasby, P. M., Frith, C. D., Friston, K. J., Bench, C., Frackowiak, R. S. J., & Dolan, R. J. (1993). Functional mapping of brain areas implicated in auditory-verbal memory function. *Brain, 116*(1), 1-20. doi:10.1093/brain/116.1.1

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468-484. doi:10.1016/0013-4694(83)90135-9

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95-112. doi:10.1007/BF02289823

Hayama, H. R., Johnson, J. D., & Rugg, M. D. (2008). The relationship between the right frontal old/new ERP effect and post-retrieval monitoring: specific or non-specific? *Neuropsychologia, 46*(5), 1211-1223. doi:10.1016/j.neuropsychologia.2007.11.021

Herron, J. E., & Wilding, E. L. (2004). An electrophysiological dissociation of retrieval mode and retrieval orientation. *Neuroimage, 22*(4), 1554-1562. doi:10.1016/j.neuroimage.2004.04.011

Herron, J. E., & Wilding, E. L. (2006a). Brain and behavioral indices of retrieval mode. *Neuroimage, 32*(2), 863-870. doi:10.1016/j.neuroimage.2006.03.046

Herron, J. E., & Wilding, E. L. (2006b). Neural correlates of control processes engaged before and during recovery of information from episodic memory. *Neuroimage, 30*(2), 634-644. doi:10.1016/j.neuroimage.2005.10.003

Hicks, J. L., & Marsh, R. L. (1998). A decrement-to-familiarity interpretation of the revelation effect from forced-choice tests of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(5), 1105-1120. doi:10.1037/0278-7393.24.5.1105

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124. doi:10.1371/journal.pmed.0020124

Jackson, A. F., & Bolger, D. J. (2014). The neurophysiological bases of EEG and EEG measurement: a review for the rest of us. *Psychophysiology, 51*(11), 1061-1071. doi:10.1111/psyp.12283

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

Janowsky, J. S., Shimamura, A. P., Kritchevsky, M., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience, 103*(3), 548-560. doi:10.1037/0735-7044.103.3.548

161

Jasper, H. H. (1958). Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology, 10*(2), 370-375. doi:10.1016/0013-4694(58)90053-1

Jeffreys, H. (1961). *Theory of probability* (3 ed.). Oxford, UK: Oxford University Press.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology, 14*(89).

Jetter, W., Poser, U., Freeman Jr, R. B., & Markowitsch, H. J. (1986). A verbal long term memory deficit in frontal lobe damaged patients. *Cortex, 22*(2), 229-242. doi:10.1016/S0010-9452(86)80047-8

Johnson, J. D., & Rugg, M. D. (2006). Electrophysiological correlates of retrieval processing: effects of consistent versus inconsistent retrieval demands. *Journal of Cognitive Neuroscience, 18*(9), 1531-1544. doi:10.1162/jocn.2006.18.9.1531

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28. doi:10.1037/0033-2909.114.1.3

Johnson, M. K., Kounios, J., & Nolde, S. F. (1997). Electrophysiological brain activity and memory source monitoring. *NeuroReport, 8*(5), 1317-1320. doi:10.1097/00001756-199703240-00051

Jost, K., De Baene, W., Koch, I., & Brass, M. (2013). A review of the role of cue processing in task switching. *Zeitschrift für Psychologie, 221*(1), 5-14. doi:10.1027/2151-2604/a000125

Jost, K., Mayr, U., & Rösler, F. (2008). Is task switching nothing but cue priming? Evidence from ERPs. *Cognitive, Affective, & Behavioral Neuroscience, 8*(1), 74-84. doi:10.3758/CABN.8.1.74

Kapur, N., Friston, K. J., Young, A., Frith, C. D., & Frackowiak, R. S. J. (1995). Activation of human hippocampal formation during memory for faces: a PET study. *Cortex, 31*(1), 99-108. doi:10.1016/S0010-9452(13)80108-6

Kapur, S., Craik, F. I. M., Jones, C., Brown, G. M., Houle, S., & Tulving, E. (1995). Functional role of the prefrontal cortex in retrieval of memories: a PET study. *NeuroReport, 6*(14), 1880-1884. doi:10.1097/00001756-199510020-00014

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., . . . Yee, C. M. (2014). Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology, 51*(1), 1-21. doi:10.1111/psyp.12147

Keppel, G. (1991). *Design and analysis: a researcher's handbook*. Englewood Cliffs, NJ: Prentice-Hall.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136*(5), 849-874. doi:10.1037/a0019842

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology, 124*(10), 2062-2063. doi:10.1016/j.clinph.2013.03.024

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*(863), 1-12. doi:10.3389/fpsyg.2013.00863

Lane, D. M., & Dunlap, W. P. (1978). Estimating effect size: bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology, 31*(2), 107-112. doi:10.1111/j.2044-8317.1978.tb00578.x

Lepage, M., Ghaffar, O., Nyberg, L., & Tulving, E. (2000). Prefrontal cortex and episodic memory retrieval mode. *Proceedings of the National Academy of Sciences, 97*(1), 506-511. doi:10.1073/pnas.97.1.506

Lepage, M., Habib, R., & Tulving, E. (1998). Hippocampal PET activations of memory encoding and retrieval: the HIPER model. *Hippocampus, 8*(4), 313-322. doi:10.1002/(SICI)1098-1063(1998)8:4<313::AID-HIPO1>3.0.CO;2-I

Liu, C. C., & Aitkin, M. (2008). Bayes factors: prior sensitivity and model generalizability. *Journal of Mathematical Psychology, 52*(6), 362-375. doi:10.1016/j.jmp.2008.03.002

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance, 29*(3), 575-599. doi:10.1037/0096-1523.29.3.575

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Mandler, G. (1980). Recognizing: the judgment of previous occurrence. *Psychological Review, 87*(3), 252-271. doi:10.1037/0033-295X.87.3.252

Marí-Beffa, P., Cooper, S., & Houghton, G. (2011). Unmixing the mixing cost: contributions from dimensional relevance and stimulus-response suppression. *Journal of Experimental Psychology: Human Perception and Performance, 38*(2), 478-488. doi:10.1037/a0025979

Mayr, U. (2001). Age differences in the selection of mental sets: the role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology and Aging, 16*(1), 96-109. doi:10.1037/0882-7974.16.1.96

Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1124-1140. doi:10.1037//0278-7393.26.5.1124

McClelland, J. L. (1989). Parallel distributed processing: implications for cognition and development. In R. G. M. Morris (Ed.), *Parallel distributed processing: implications for psychology and neurobiology* (pp. 8-45). Oxford, UK: Clarendon Press.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*(2), 361-365. doi:10.1037/0033-2909.111.2.361

Mecklinger, A. (2010). The control of long-term memory: brain systems and cognitive processes. *Neuroscience and Biobehavioral Reviews, 34*(7), 1055-1065. doi:10.1016/j.neubiorev.2009.11.020

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology, 41*(3), 211-253. doi:10.1006/cogp.2000.0736

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134-140. doi:10.1016/S1364-6613(03)00028-7

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition, 31*(3), 327-342. doi:10.3758/BF03194391

Morcom, A. M., & Rugg, M. D. (2002). Getting ready to remember: the neural correlates of task set during recognition memory. *NeuroReport, 13*(1), 149-152. doi:10.1097/00001756-200201210-00034

Morey, R. D., Rouder, J. N., Verhagen, J., & Wagenmakers, E.-J. (2014). Why hypothesis tests are essential for psychological science: a comment on Cumming (2014). *Psychological Science, 25*(6), 1289-1290. doi:10.1177/0956797614525969

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*(5), 519-533. doi:10.1016/S0022-5371(77)80016-9

Moscovitch, M. (1989). Confabulation and the frontal systems: strategic versus associative retrieval in neuropsychological theories of memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: essays in honour of Endel Tulving* (pp. 133-160). Hillsdale, NJ: Lawrence Erlbaum Associates.

Moscovitch, M. (1992). Memory and working-with-memory: a component process model based on modules and central systems. *Journal of Cognitive Neuroscience, 4*(3), 257-267. doi:10.1162/jocn.1992.4.3.257

Nicholson, R., Karayanidis, F., Bumak, E., Poboka, D., & Michie, P. T. (2006). ERPs dissociate the effects of switching task sets and task cues. *Brain Research, 1095*(1), 107-123. doi:10.1016/j.brainres.2006.04.016

Norman, D. A., & Bobrow, D. G. (1979). Descriptions: an intermediate stage in memory retrieval. *Cognitive Psychology, 11*(1), 107-123. doi:10.1016/0010-0285(79)90006-9

Norman, D. A., & Shallice, T. (1986). Attention to action: willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self regulation: advances in research* (Vol. 4, pp. 1-18). New York, NY: Plenum Press.

Nunez, P., & Srinivasan, R. (1981). *Electrical fields of the brain*. New York, NY: Oxford University Press.

Nyberg, L., Cabeza, R., & Tulving, E. (1996). PET studies of encoding and retrieval: the HERA model. *Psychonomic Bulletin & Review, 3*(2), 135-148. doi:10.3758/BF03212412

Nyberg, L., Persson, J., Habib, R., Tulving, E., McIntosh, A. R., Cabeza, R., & Houle, S. (2000). Large scale neurocognitive networks underlying episodic memory. *Journal of Cognitive Neuroscience, 12*(1), 163-173. doi:10.1162/089892900561805

Nyberg, L., Tulving, E., Habib, R., Nilsson, L.-G., Kapur, S., Houle, S., . . . McIntosh, A. R. (1995). Functional brain maps of retrieval mode and recovery of episodic information. *NeuroReport, 7*(1), 249-252. doi:35400005289013.0580

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods, 8*(4), 434-447. doi:10.1037/1082-989X.8.4.434

Otten, L. J., Henson, R. N. A., & Rugg, M. D. (2002). State-related and item-related neural correlates of successful memory encoding. *Nature Neuroscience, 5*(12), 1339-1344. doi:10.1038/nn967

Otten, L. J., Quayle, A. H., Akram, S., Ditewig, T. A., & Rugg, M. D. (2006). Brain activity before an event predicts later recollection. *Nature Neuroscience, 9*(4), 489-491. doi:10.1038/nn1663

Otten, L. J., Quayle, A. H., & Puvaneswaran, B. (2010). Prestimulus subsequent memory effects for auditory and visual events. *Journal of Cognitive Neuroscience, 22*(6), 1212-1223. doi:10.1162/jocn.2009.21298

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., . . . Rugg, M. D. (2000). Guidelines for using human event-related potentials to study cognition:

recording standards and publication criteria. *Psychophysiology, 37*(2), 127-152. doi:10.1017/S0048577200000305

Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N., & Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology, 30*(6), 547-558. doi:10.1111/j.1469-8986.1993.tb02081.x

Poljac, E., Koch, I., & Bekkering, H. (2009). Dissociating restart cost and mixing cost in task switching. *Psychological Research, 73*(3), 407-416. doi:10.1007/s00426-008-0151-9

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological Methods, 16*(2), 93-115. doi:10.1037/a0022658

Ranganath, C., & Knight, R. T. (2002). Prefrontal cortex and episodic memory: integrating findings from neuropsychology and functional brain imaging. In A. Parker, E. L. Wilding, & T. J. Bussey (Eds.), *The cognitive neuroscience of memory: encoding and retrieval* (Vol. 1, pp. 83-99). New York, NY: Psychology Press.

Ranganath, C., & Paller, K. A. (1999). Frontal brain potentials during recognition are modulated by requirements to retrieve perceptual detail. *Neuron, 22*(3), 605-613. doi:10.1016/S0896-6273(00)80714-X

Reinitz, M. T., Lammers, W. J., & Cochran, B. P. (1992). Memory-conjunction errors: miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition, 20*(1), 1-11. doi:10.3758/BF03208247

Robb, W. G. K., & Rugg, M. D. (2002). Electrophysiological dissociation of retrieval orientation and retrieval effort. *Psychonomic Bulletin & Review, 9*(3), 583-589. doi:10.3758/BF03196316

Rockstroh, B., Elbert, T., Canavan, A., Lutzenberger, W., & Birbaumer, N. (1989). *Slow cortical potentials and behavior*. Baltimore, MD: Urban and Schwarzenberg.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207-230. doi:10.1037/0096-3445.124.2.207

Rosburg, T., Mecklinger, A., & Johansson, M. (2011). Electrophysiological correlates of retrieval orientation in reality monitoring. *Neuroimage, 54*(4), 3076-3084. doi:10.1016/j.neuroimage.2010.10.068

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641. doi:10.1037/0033-2909.86.3.638

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.

Rotello, C. M., Masson, M. E. J., & Verde, M. F. (2008). Type I error rates and power analyses for single-point sensitivity measures. *Perception & Psychophysics, 70*(2), 389-401. doi:10.3758/PP.70.2.389

Rouder, J. N. (2014). Optional stopping: no problem for Bayesians. *Psychonomic Bulletin & Review, 21*(2), 301-308. doi:10.3758/s13423-014-0595-4

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review, 18*(4), 682-689. doi:10.3758/s13423-011-0088-7

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356-374. doi:10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225-237. doi:10.3758/PBR.16.2.225

Rubin, O., & Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1477-1491. doi:10.1037/0278-7393.31.6.1477

Rugg, M. D. (1995). ERP studies of memory. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: event-related potentials and cognition* (pp. 132-170). New York, NY: Oxford University Press.

Rugg, M. D., Allan, K., & Birch, C. S. (2000). Electrophysiological evidence for the modulation of retrieval orientation by depth of study processing. *Journal of Cognitive Neuroscience, 12*(4), 664-678. doi:10.1162/089892900562291

Rugg, M. D., & Coles, M. G. H. (1995). *Electrophysiology of mind: event-related brain potentials and cognition*. New York, NY: Oxford University Press.

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

Rugg, M. D., Fletcher, P. C., Allan, K., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1998). Neural correlates of memory retrieval during recognition memory and cued recall. *Neuroimage, 8*(3), 262-273. doi:10.1006/nimg.1998.0363

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1996). Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain, 119*(6), 2073-2083. doi:10.1093/brain/119.6.2073

Rugg, M. D., & Henson, R. N. A. (2002). Episodic memory retrieval: an (event-related) functional neuroimaging perspective. In A. Parker, E. L. Wilding, & T. J. Bussey (Eds.), *The cognitive neuroscience of memory: encoding and retrieval* (Vol. 1, pp. 3-37). New York, NY: Psychology Press.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences, 4*(3), 108-115. doi:10.1016/S1364-6613(00)01445-5

Salthouse, T. A., Fristoe, N., McGuthry, K. E., & Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. *Psychology and Aging, 13*(3), 445-461. doi:10.1037/0882-7974.13.3.445

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., & Rosen, B. R. (1997). Late onset of anterior prefrontal activity during true and false recognition: an event-related fMRI study. *Neuroimage, 6*(4), 259-269. doi:10.1006/nimg.1997.0305

Schacter, D. L., Eich, J. E., & Tulving, E. (1978). Richard Semon's theory of memory. *Journal of Verbal Learning and Verbal Behavior, 17*(6), 721-743. doi:10.1016/S0022-5371(78)90443-7

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49*(1), 289-318. doi:10.1146/annurev.psych.49.1.289

Schacter, D. L., Reiman, E., Curran, T., Yun, L. S., Bandy, D., McDermott, K. B., & Roediger III, H. L. (1996). Neuroanatomical correlates of veridical and illusory recognition memory: evidence from positron emission tomography. *Neuron, 17*(2), 267-274. doi:10.1016/S0896-6273(00)80158-0

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66. doi:10.1037/0033-295X.84.1.1

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of ρ values for testing precise null hypotheses. *The American Statistician, 55*(1), 62-71. doi:10.1198/000313001300339950

Semon, R. (1921). *The mneme (L. Simon, Trans.).* London, UK: George Allen & Unwin. (Original work published 1904).

Shallice, T., Fletcher, P., Frith, C. D., Grasby, P., Frackowiak, R. S. J., & Dolan, R. J. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature, 368*(6472), 633-635. doi:10.1038/368633a0

Sharbrough, F., Chatrian, G. E., Lesser, R. P., Lüders, H., Nuwer, M., & Picton, T. W. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology, 8*(2), 200-202.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366. doi:10.1177/0956797611417632

Simons, J. S., Gilbert, S. J., Owen, A. M., Fletcher, P. C., & Burgess, P. W. (2005). Distinct roles for lateral and medial anterior prefrontal cortex in contextual recollection. *Journal of Neurophysiology, 94*(1), 813-820. doi:10.1152/jn.01200.2004

Simons, J. S., Owen, A. M., Fletcher, P. C., & Burgess, P. W. (2005). Anterior prefrontal cortex and the recollection of contextual information. *Neuropsychologia, 43*(12), 1774-1783. doi:10.1016/j.neuropsychologia.2005.02.004

Smith, M. E. (1993). Neurophysiological manifestations of recollective experience during recognition memory judgments. *Journal of Cognitive Neuroscience, 5*(1), 1-13. doi:10.1162/jocn.1993.5.1.1

Smolensky, P. (1987). The constituent structure of connectionist mental states: a reply to Fodor and Pylyshyn. *The Southern Journal of Philosophy, 26*(S1), 137-161. doi:10.1111/j.2041-6962.1988.tb00470.x

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. doi:10.1037/0096-3445.117.1.34

Squire, L. R., Ojemann, J. G., Miezin, F. M., Petersen, S. E., Videen, T. O., & Raichle, M. E. (1992). Activation of the hippocampus in normal humans: a functional anatomical study of memory. *Proceedings of the National Academy of Sciences, 89*(5), 1837-1841. doi:10.1073/pnas.89.5.1837

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662. doi:10.1037/h0054651

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., . . . Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language, 51*(2), 247-250. doi:10.1016/j.jml.2004.03.002

169

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain.* New York, NY: Thieme Medical.

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools, 44*(5), 423-432. doi:10.1002/pits.20234

Toga, A. W., & Mazziotta, J. C. (2002). *Brain mapping: the methods* (2 ed.). San Diego, CA: Academic Press.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381-402). New York, NY: Acdemic Press.

Tulving, E. (1983). *Elements of episodic memory*. Oxford, UK: Clarendon Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*(1), 1-12. doi:10.1037/h0080017

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science, 2*(3), 67-70. doi:10.1111/1467-8721.ep10770899

Tulving, E., Kapur, S., Craik, F. I., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings. *Proceedings of the National Academy of Sciences, 91*(6), 2016-2020. doi:10.1073/pnas.91.6.2016

Tulving, E., Kapur, S., Markowitsch, H. J., Craik, F. I., Habib, R., & Houle, S. (1994). Neuroanatomical correlates of retrieval in episodic memory: auditory sentence recognition. *Proceedings of the National Academy of Sciences, 91*(6), 2012-2015. doi:10.1073/pnas.91.6.2012

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 381-391. doi:10.1016/S0022-5371(66)80048-8

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352-373. doi:10.1037/h0020071

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: interplay of reconfiguration and interference control. *Psychological Bulletin, 136*(4), 601-626. doi:10.1037/a0019791

Vanpaemel, W. (2010). Prior sensitivity in theory testing: an apologia for the Bayes factor. *Journal of Mathematical Psychology, 54*(6), 491-498. doi:10.1016/j.jmp.2010.07.003

Velanova, K., Jacoby, L. L., Wheeler, M. E., McAvoy, M. P., Petersen, S. E., & Buckner, R. L. (2003). Functional-anatomic correlates of sustained and transient processing

components engaged during controlled retrieval. *The Journal of Neuroscience, 23*(24), 8460-8470. doi:0270-6474/03/238460-11$15.00/0

Verde, M. F., & Rotello, C. M. (2003). Does familiarity change in the revelation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(5), 739-746. doi:10.1037/0278-7393.29.5.739

Verde, M. F., & Rotello, C. M. (2004). ROC curves show that the revelation effect is not a single phenomenon. *Psychonomic Bulletin & Review, 11*(3), 560-566. doi:10.3758/BF03196611

Verhagen, A. J., & Wagenmakers, E.-J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General, 143*(4), 1457-1475. doi:10.1037/a0036731

Watkins, M. J., & Peynircioglu, Z. F. (1990). The revelation effect: when disguising test items induces recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(6), 1012-1020. doi:10.1037/0278-7393.16.6.1012

Werkle-Bergner, M., Mecklinger, A., Kray, J., Meyer, P., & Düzel, E. (2005). The control of memory retrieval: insights from event-related potentials. *Cognitive Brain Research, 24*(3), 599-614. doi:10.1016/j.cogbrainres.2005.03.011

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology an empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6*(3), 291-298. doi:10.1177/1745691611406923

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1995). Frontal lobe damage produces episodic memory impairment. *Journal of the International Neuropsychological Society, 1*(6), 525-536. doi:10.1017/S1355617700000655

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychological Bulletin, 121*(3), 331-354. doi:10.1037/0033-2909.121.3.331

Wilckens, K. A., Tremel, J. J., Wolk, D. A., & Wheeler, M. E. (2011). Effects of task-set adoption on ERP correlates of controlled and automatic recognition memory. *Neuroimage, 55*(3), 1384-1392. doi:10.1016/j.neuroimage.2010.12.059

Wilding, E. L. (1999). Separating retrieval strategies from retrieval success: an event-related potential study of source memory. *Neuropsychologia, 37*(4), 441-454. doi:10.1016/s0028-3932(98)00100-6

171

Wilding, E. L. (2001). Event-related functional imaging and episodic memory. *Neuroscience & Biobehavioral Reviews, 25*(6), 545-554. doi:10.1016/S0149-7634(01)00033-1

Wilding, E. L., Doyle, M. C., & Rugg, M. D. (1995). Recognition memory with and without retrieval of context: an event-related potential study. *Neuropsychologia, 33*(6), 743-767. doi:10.1016/0028-3932(95)00017-W

Wilding, E. L., & Herron, J. E. (2006). Electrophysiological measures of episodic memory control and memory retrieval. *Clinical EEG and Neuroscience, 37*(4), 315-321. doi:10.1177/155005940603700409

Wilding, E. L., & Nobre, A. C. (2001). Task-switching and memory retrieval processing: electrophysiological evidence. *NeuroReport, 12*(16), 3613-3617. doi:10.1097/00001756-200111160-00048

Wilding, E. L., & Ranganath, C. (2011). Electrophysiological correlates of episodic memory processes. In S. J. Luck & E. S. Kappenman (Eds.), *The oxford handbook of event-related potential components* (pp. 373-395). New York, NY: Oxford University Press.

Wilding, E. L., & Rugg, M. D. (1996). An event-related potential study of recognition memory with and without retrieval of source. *Brain, 119*(3), 889-905. doi:10.1093/brain/119.3.889

Wilding, E. L., & Sharpe, H. (2003). Episodic memory encoding and retrieval: recent insights from event-related potentials. In A. Zani & A. Mado-Proverbio (Eds.), *The cognitive electrophysiology of mind and brain* (pp. 169-196). San Diego, CA: Academic Press.

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

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychological Research, 63*(3-4), 212-233. doi:10.1007/s004269900003

Yeung, N., Nystrom, L. E., Aronson, J. A., & Cohen, J. D. (2006). Between-task competition and cognitive control in task switching. *The Journal of Neuroscience, 26*(5), 1429-1438. doi:10.1523/JNEUROSCI.3109-05.2006

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., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: effects of size congruency. *Journal of Memory and Language, 34*(5), 622-643. doi:10.1006/jmla.1995.1028

10. Appendices

Appendix A. Effect size calculations and *a priori* power analyses

Appendix B. Test item ERP data: Experiments 1-4

Appendix C. Replication of Evans, Williams, et al. (2015): Experiments 1-4

Appendix D. Behavioural results summary table: Experiments 1-6

10.1 Appendix A. Effect size calculations and a priori power analyses

10.1.1 Experiment 1

Effect sizes were calculated for the preparatory ERP effects of interest reported in previous studies, namely more positive-going activity at right-frontal scalp sites when preparing for episodic rather than non-episodic tasks (**Table A.1**; see **General methods** for effect size equations). These effect sizes were then used in *a priori* power analyses, to estimate how many participants would be required for a replication attempt (**Table A.2**). 'CL' stands for the Common Language effect size. This effect size aids intuitive interpretation, as it simply reflects the probability that an individual participant has a greater value on one measurement than another (Lakens, 2013; McGraw & Wong, 1992).

Table A.1. Effect sizes for preparatory ERP effects in key studies.

Study	**Effect**	*N*	*F* **statistics**	*t* **value**	**Effect size**
Düzel et al. (1999)	Main effect of task, blocked study	11	$F(1, 10) = 5.2$, $p < 0.05$	2.28	Cohen's d_z = 0.69
Morcom & Rugg (2002)	Main effect of task, switching study	20	$F(1, 19) = 4.93$, $p < 0.05$	2.22	Cohen's d_z = 0.50
	Task by site interaction, stay trials		$F(1.8, 33.5) = 5.88$, $p < 0.01$	N/A	Partial eta^2 = 0.24
Herron & Wilding (2004)	Main effect of task (operation v semantic, stay trials)	20	$F(1, 19) = 4.37$, $p < 0.05$	2.09	Cohen's d_z = 0.47
	Main effect of task (location v semantic, stay trials)		$F(1, 19) = 8.20$, $p < 0.01$	2.86	Cohen's d_z = 0.64
Herron & Wilding (2006a)	Main effect of task (location v semantic, stay trials)	16	$F(1, 15) = 6.08$, $p < 0.05$	2.47	Cohen's d_z = 0.62
Evans, Williams, et al. (2015)	Main effect of task (location v perceptual, switch trials)	32	$F(1, 31) = 9.54$, $p < 0.01$	3.09	Cohen's d_z = 0.55

Table A.2. Results of *a priori* power analyses (N_{req}) for replication attempts of the effects reported in Table A.1.

Study	Effect	N_{orig}	Effect size	Power $(1 - \beta)$	Alpha (α)	Test	CL	N_{req}
Düzel et al. (1999)	Main effect of task, blocked study	11	Cohen's d_z = 0.69	0.80	0.05	One-sided t	75%	15
Morcom & Rugg (2002)	Main effect of task, switching study	20	Cohen's d_z = 0.50	0.80	0.05	One-sided t	69%	27
	Task by site interaction, stay trials*		Partial eta^2 = 0.24	0.80	0.05	ANOVA	N/A	20
Herron & Wilding (2004)	Main effect of task (operation v semantic, stay trials)	20	Cohen's d_z = 0.47	0.80	0.05	One-sided t	69%	30
	Main effect of task (location v semantic, stay trials)		Cohen's d_z = 0.64	0.80	0.05	One-sided t	74%	17
Herron & Wilding (2006a)	Main effect of task (location v semantic, stay trials)	16	Cohen's d_z = 0.62	0.80	0.05	One-sided t	73%	18
Evans, Williams, et al. (2015)	Main effect of task (location v perceptual, switch trials)	32	Cohen's d_z = 0.55	0.80	0.05	One-sided t	71%	22

*Additional parameters are required as input for power calculations of interaction effects. For example, for the Morcom and Rugg (2002) effect: the number of groups, measurements, and the Nonsphericity correction €, were required. In addition, if the effect was corrected for sphericity, with corrected dfs used accordingly, then G*Power3.1.7 requires that the default 'Options' are reset according to SPSS (Faul el at., 2007; Lakens, 2013). Here, the number of groups was: 2 (tasks), the number of measurements was: 4 (clusters of sites), and the € was: 0.6 (calculated by working out the uncorrected df numerator = 4-1 = 3, thus, 1.8 (corrected df) /3 = 0.6).

For the experiments reported in this thesis, a maximum sample size of 32 participants was adopted. This was based on the maximum sample size used previously (**Table A.1**), and exceeds all the required sample sizes according to the *a priori* power analyses for replication of the right-frontally distributed preparatory ERP indices reported in the literature (**Table A.2**).

On average, the sample size required (N_{req}) to replicate the more positive-going activity at right-frontal scalp sites when preparing for episodic in comparison to non-episodic tasks (evident in previous studies using a task-switching design) is 22 participants. For Experiment 1, a sample size of 24 participants was predetermined based on these *a priori* power analyses, and counterbalancing constraints.

Following successful data collection from 24 participants, *F/t* statistics were obtained for the effects of interest, and subsequent Bayesian analyses were carried out. The Bayes Factors (BFs) for the right-frontally distributed preparatory ERP index provided strong evidence in favour of the null hypothesis ($BF_{ave} = 0.07$). Thus, data collection was terminated at this point, in line with the Bayesian Stopping Rule (Dienes, 2011; Verhagen & Wagenmakers, 2014).

10.1.2 Experiment 2

Effect sizes were calculated for the effects evident in Experiment 1 (behavioural and ERP; **Table A.3**). These effect sizes were then used in *a priori* power analyses, to estimate how many participants would be required for a replication attempt (**Table A.4**).

The average sample size required (N_{req}) to replicate the effects evident in Experiment 1 is 29 participants. However, for Experiment 2 data were initially collected from 24 participants, in accordance with the *a priori* power analyses for the right-frontally distributed preparatory ERP index and counterbalancing demands. Again, *F/t* statistics were obtained for the effects of interest, and subsequent Bayesian analyses were carried out. At 24 participants there was evidence for a behavioural old/new accuracy switch cost, however, the BFs for the right-frontally distributed preparatory ERP index did not provide evidence in favour of the null nor alternative hypothesis. Thus, further data collection was necessary, and an additional eight datasets were collected, which resulted in a fully counterbalanced dataset with 32 participants.

Table A.3. Effect sizes for the effects evident in Experiment 1.

Study	Effect	N	*F/t* statistics	Effect size
Experiment 1	Old/new accuracy switch cost	24	$t(23) = 2.56$, $p < 0.05$	Cohen's d_z $= 0.52$
	Source accuracy switch cost		$t(23) = 2.77$, $p < 0.05$	Cohen's d_z $= 0.57$
	Cue-type by hemisphere interaction, switch trials anterior sites		$F(1, 23) = 8.31$ $p < 0.05$	Partial eta^2 $= 0.27$
	Cue-type by site interaction, switch trials anterior sites		$F(1.3, 29.6) = 4.00$, $p < 0.05$	Partial eta^2 $= 0.15$

Table A.4. Results of *a priori* power analyses (N_{req}) for replication attempts of the results in Experiment 1.

Study	Effect	N_{orig}	Effect size	Power	Alpha (α)	Test	CL	N_{req}
Experiment 1	Old/new accuracy switch cost	24	Cohen's d_z $= 0.52$	0.80	0.05	One-sided t	70%	25
	Source accuracy switch cost		Cohen's d_z $= 0.57$	0.80	0.05	One-sided t	71%	21
	Cue-type by hemisphere interaction, switch trials anterior sites		Partial eta^2 $= 0.27$	0.80	0.05	ANOVA	N/A	26
	Cue-type by site interaction, switch trials anterior sites		Partial eta^2 $= 0.15$	0.80	0.05	ANOVA	N/A	42

On average, the BFs calculated from the results from 32 participants provided anecdotal evidence in favour of the null hypothesis (BF_{ave} = 0.55), and data collection was terminated at this point. Increasing data collection from 24 to 32 participants moved the BFs from providing 'no evidence' towards providing 'anecdotal and substantial evidence in favour of the null hypothesis'.

10.1.3 Experiment 3

A sample size of 24 participants was initially determined for Experiment 3. This was based on the *a priori* power analyses from the previous studies identifying the right-frontally distributed preparatory ERP index using a task-switching design, and on counterbalancing demands.

Following successful data collection from 24 participants, *F/t* statistics were obtained for the effects of interest, and subsequent Bayesian analyses were carried out. The BFs revealed anecdotal evidence for a behavioural reaction time switch cost in favour of the alternative hypothesis, when traditional NHST did not provide support for an effect. Thus, further data collection was necessary. A further eight datasets were collected, completing a fully counterbalanced dataset with 32 participants.

Subsequent analyses from the 32 datasets revealed: a reaction time switch cost, and an anecdotal BF in support of the alternative hypothesis for an old/new discrimination accuracy switch cost. BF analyses were also carried out on the ERP results from 32 datasets, providing substantial evidence in favour of the null hypothesis (BF_{ave} = 0.33). Thus, data collection was terminated at this point.

10.1.4 Experiment 4

A sample size of 24 participants was initially determined for Experiment 4. This was based on the *a priori* power analyses from the previous studies identifying the right-frontally distributed preparatory ERP index using a task-switching design, and on counterbalancing demands.

Following successful data collection from 24 participants, *F/t* statistics were obtained for the effects of interest, and subsequent Bayesian analyses were carried out. For the right-frontally distributed preparatory ERP index on stay trials, the BFs revealed support for the null hypothesis. However, as a replication of Evans, Williams, et al. (2015) (see **General discussion**) for the right-frontally distributed preparatory ERP index on switch trials, the BF provided anecdotal evidence for the alternative hypothesis. Thus, further data collection was necessary. A further eight datasets were collected, completing a fully counterbalancing dataset with 32 participants.

The BFs calculated from the results of 32 participants also provided anecdotal support of the alternative hypothesis for: an old/new discrimination accuracy switch cost, and the right-frontally distributed preparatory ERP index on switch trials.

10.1.5 Experiment 5

The average sample size required (N_{req}) to replicate the behavioural effects evident in Experiment 1, was 23 participants. For Experiment 5, 24 datasets were collected (based on the *a priori* power analyses and counterbalancing demands). *F/t* statistics were obtained for the effects of interest, and subsequent Bayesian analyses were carried out. The BFs provided very strong support for the alternative hypothesis for an old/new discrimination accuracy switch cost (switch < stay), and substantial support for the null hypothesis for the source accuracy switch cost. Thus, data collection was terminated at this point.

10.1.6 Experiment 6

In line with Experiments 1 and 5, 24 datasets were collected (based on counterbalancing demands). More detailed *a priori* power analyses were not required for this experiment, as the design differed more substantially from that employed in previous experiments. *F/t* statistics were obtained for the effects of interest, and subsequent Bayesian analyses were carried out. The BFs provided very strong support for the alternative hypothesis for an old/new discrimination accuracy switch cost (switch < stay), and strong support for the null hypothesis for the source accuracy switch cost. Thus, data collection was terminated at this point.

179

In reports of investigations of retrieval mode, discussions have included the possible function and benefits of adopting retrieval sets. Herron and Wilding (2004) noted that the benefits of adopting retrieval sets can be investigated by analysing ERP indices of processes related to subsequent retrieval success (old/new effects). Examination of the ways in which the old/new effects vary for switch and stay trials offers a way to assess the benefits successful adoption has for subsequent retrieval processing.

One way in which successful adoption could influence subsequent processing, for example, is by increasing the quality or volume of content that is retrieved. In keeping with this account, Evans et al. (2012) reported larger left-parietal ERP old/new effects on stay+1 than on stay and switch trials, which is consistent with the view that when the time available to adopt a set is increased the quality or volume of recovered episodic information also increases (see also Wilckens et al., 2011). Interestingly, Evans, Herron, et al. (2015) reported what they described as a larger index of recollection on switch than on stay trials for their non-episodic (perceptual) task via a contrast of ERP repetition effects, where data are separated according to old/new status. They argued that this outcome indicated that processing relevant to the preceding (episodic) task carried over to the perceptual task, and carried over to a greater degree on switch than on stay trials. They argued that this outcome provided support for the concept of task-set inertia, whereby switch costs are due, at least in part, to interference from the task completed on the preceding trial (Allport et al., 1994; Wylie & Allport, 2000).

As the investigation of ERPs elicited by test items was not the primary aim of the experiments contained in this thesis, the experiments were constructed to maximise the trial numbers for the preparatory cue data whilst considering the duration of the task to aid participant motivation and avoid fatigue. As a result of this emphasis, in the four ERP experiments reported in this thesis contrasts between the ERPs elicited by test items separated by old/new status (and old/new accuracy for the episodic task) were only possible with subsets of participants for whom sufficient trial numbers were available.

For the same reasons, for the episodic tasks, old/new effects were not separated according to source accuracy. In accordance with the approach taken in previous ERP studies

in which recognition memory was examined (for reviews see Friedman & Johnson, 2000; Wilding & Sharpe, 2003), the differences between old and new stimuli attracting correct judgments were analysed in a 500-800ms post-stimulus epoch. This time window maps onto the oft-reported left-parietal ERP old/new effect, as mentioned in the **General introduction** to this thesis, which is proposed to index recollection (Jacoby, 1991; Mandler, 1980; Woodruff et al., 2006; Yonelinas, 2002). The same time window was employed for the analysis of repetition effects in the non-episodic tasks. A later time window (900-1200ms) was also explored to investigate an effect that has been linked to post-retrieval processing: the late right-frontal ERP old/new effect (Hayama et al., 2008; Wilding & Rugg, 1996).

Separate ANOVAs were conducted for the two epochs (500-800, 900-1200ms) for each of the tasks (episodic and semantic) incorporating the factors of hemisphere (left/right), site (mid-lateral/superior), trial-type (switch/stay), and the object status (semantic task: old/new, episodic task: hit old/correct rejection new). Only outcomes involving the factor of object status are reported. Each analysis included four sites: two parietal sites from each hemisphere for the 500-800ms epoch (P5/P6, P3/P4) and two frontal sites from each hemisphere for the 900-1200ms epoch (F5/F6, F3/F4).

10.2.1 Experiment 1

This analysis included the data from a subset of 15 participants where sufficient trial numbers (≥16) were available for comparisons in each of the conditions of interest. On average, 86% of the available trials contributed to the ERP test item data in this experiment for each participant. For the test item data the mean trial numbers contributing to the ERPs (ranges in parenthesis) were: episodic switch old = 23 (16-29), episodic stay old = 25 (18-30), episodic switch new = 22 (16-28), episodic stay new = 25 (21-29), semantic switch old = 28 (24-30), semantic stay old = 28 (25-30), semantic switch new = 27 (23-30), semantic stay new = 29 (24-30).

Figure B.1 demonstrates the grand averaged ERP waveforms associated with item status (old/new) for each task at a representative right anterior (F6) and left posterior (P5) site, separated for switch and stay trials. Scalp maps depicting the ERP old/new effects (episodic task) and repetition effects (semantic task) are shown for each epoch in **Figure B.2** (A. 500-800ms, B. 900-1200ms) separated for switch and stay trials.
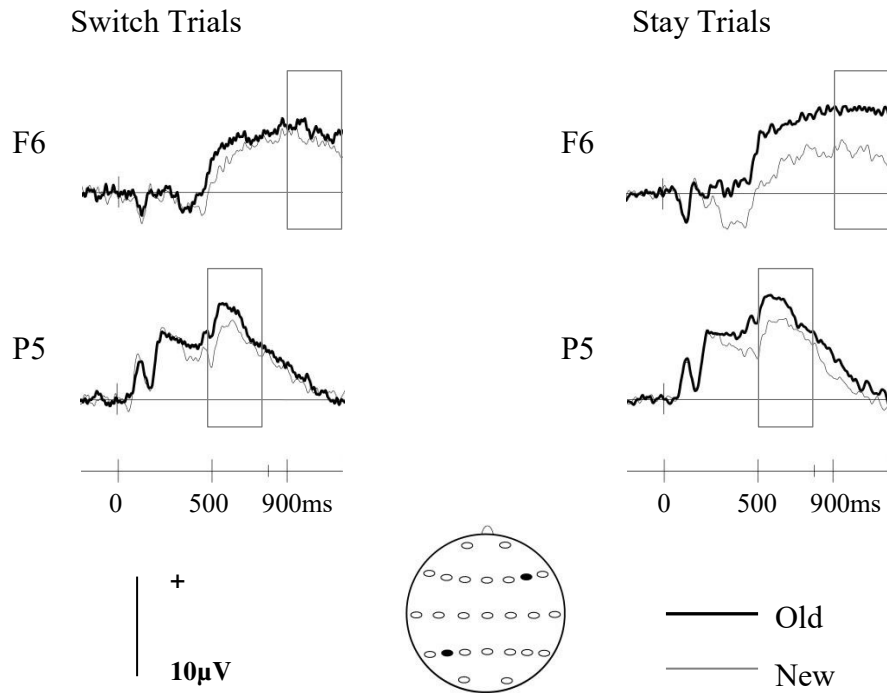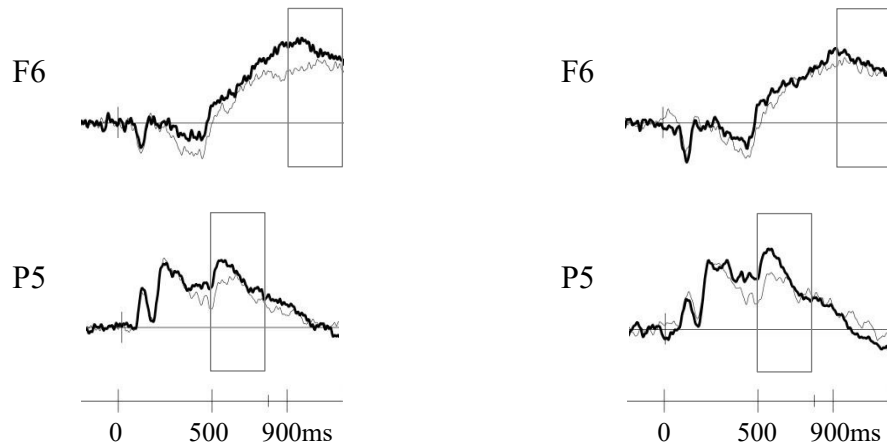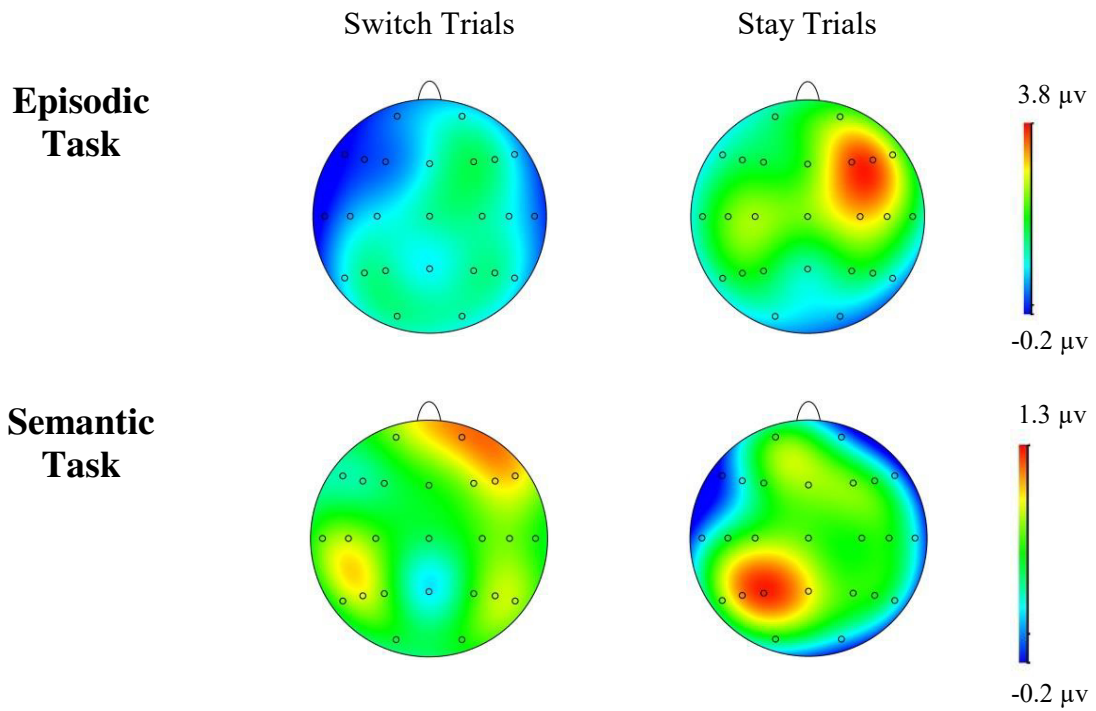
Figure B.1. Grand averaged ERPs associated with old and new test items in Experiment 1 for switch and stay trials at a representative right anterior (F6) and left posterior electrode site (P5). For the semantic task the data are separated only by old/new status. For the episodic task the data are separated for items attracting correct old or new responses.
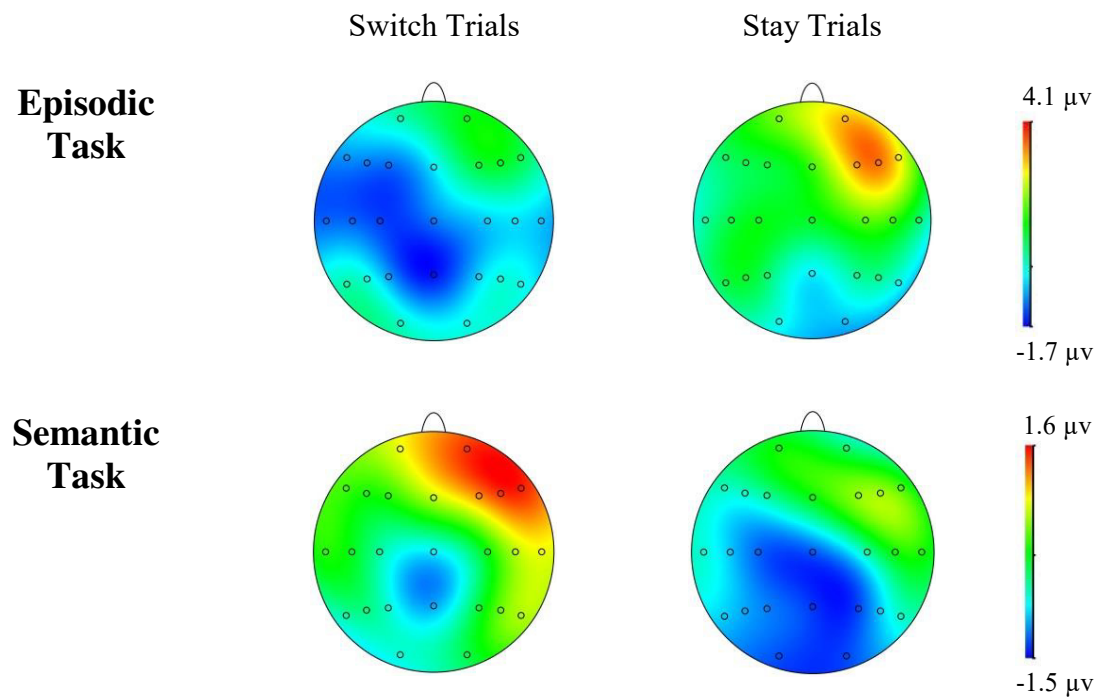
Figure B.2. Topographic maps for Experiment 1 test item data: the scalp distributions of the old/new effects (episodic task) and the repetition effects (semantic task) on switch and stay trials (*A. 500-800ms; B. 900-1200ms*).

10.2.1.1 Episodic task

In the 500-800ms epoch, there were no reliable old/new effects. In the 900-1200ms epoch, there was an interaction between old/new status and hemisphere ($F(1, 14) = 34.55$, $p < 0.001$, $\eta_p^2 = 0.71$) reflecting greater relative positivity for old than new items over the right hemisphere and greater relative positivity for new items over the left hemisphere.
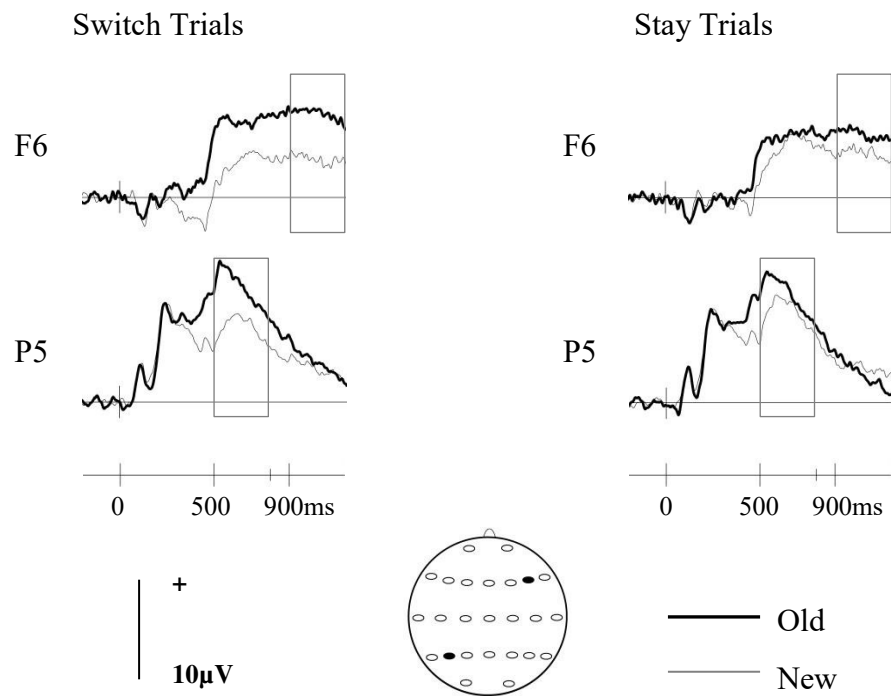
10.2.1.2 Semantic task

In the 500-800ms epoch, there was a main effect of old/new status ($F(1, 14) = 4.95$, $p < 0.05$, $d_z = 0.59$, 72% CL) moderated by an old/new by switch/stay by hemisphere by site interaction ($F(1, 14) = 5.13$, $p < 0.05$, $\eta_p^2 = 0.27$). Separate ANOVAs were conducted for switch and stay trials, and there were no reliable repetition effects. The reliable higher order interaction likely reflects the greatest relative positivity for old than new items over the left hemisphere superior site (P3) for stay trials, and then for switch trials at the mid-lateral site (P5). In the 900-1200ms epoch, there was an interaction between old/new status and hemisphere ($F(1, 14) = 13.21$, $p < 0.05$, $\eta_p^2 = 0.49$) reflecting greater relative positivity for old than new items over the right hemisphere and greater relative positivity for new items over the left hemisphere.

10.2.2 Experiment 2

This analysis included the data from a subset of 23 participants where sufficient trial numbers (≥ 16) were available for comparisons in each of the conditions of interest. On average, 88% of the available trials contributed to the ERP test item data in this experiment for each participant. For the test item data the mean trial numbers contributing to the ERPs (ranges in parenthesis) were: episodic switch old = 23 (18-29), episodic stay old = 25 (18-30), episodic switch new = 23 (16-29), episodic stay new = 26 (19-30), semantic switch old = 27 (24-30), semantic stay old = 28 (23-30), semantic switch new = 28 (24-30), semantic stay new = 29 (24-30).

Figure B.3 demonstrates the grand averaged ERP waveforms associated with item status (old/new) for each task at a representative right anterior (F6) and left posterior (P5) site, separated for switch and stay trials. Scalp maps depicting the ERP old/new effects (episodic task) and repetition effects (semantic task) are shown for each epoch in **Figure B.4** (A. 500-800ms, B. 900-1200ms) separated for switch and stay trials.
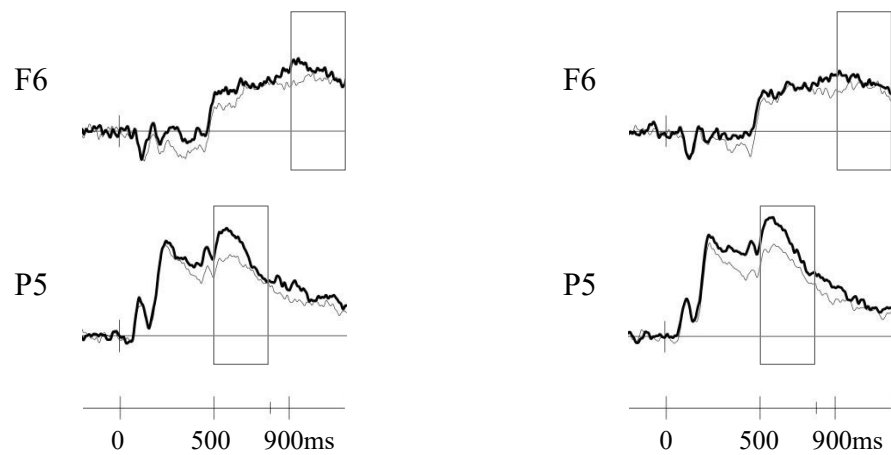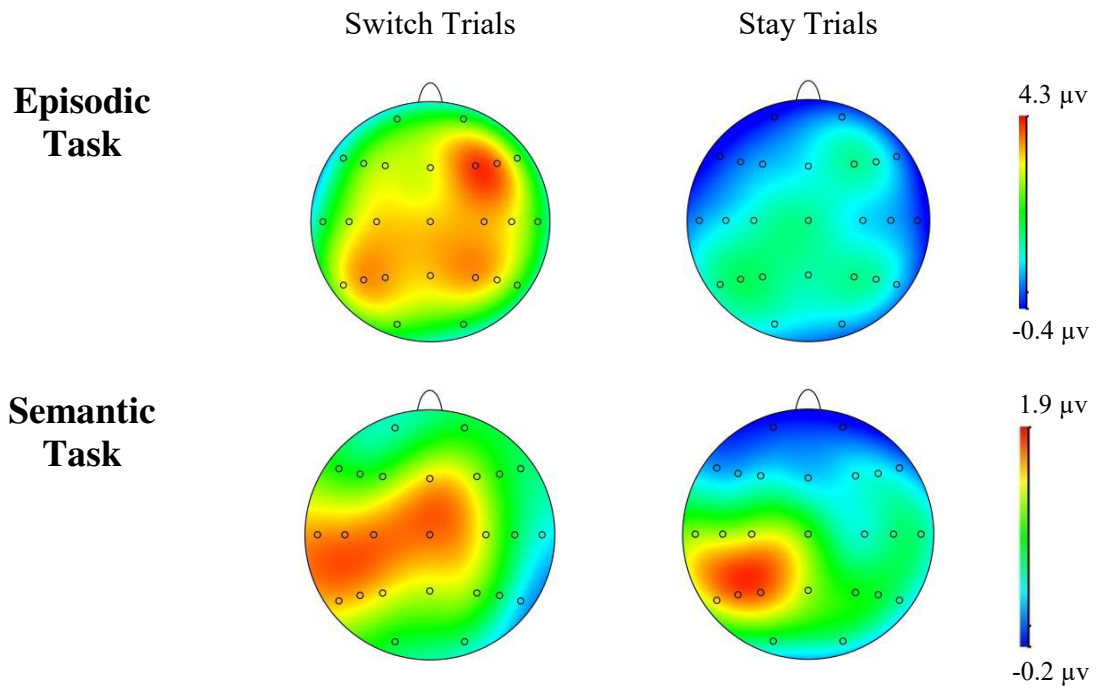
Figure B.3. Grand averaged ERPs associated with old and new test items in Experiment 2 for switch and stay trials at a representative right anterior (F6) and left posterior electrode site (P5). For the semantic task the data are separated only by old/new status. For the episodic task the data are separated for items attracting correct old or new responses.
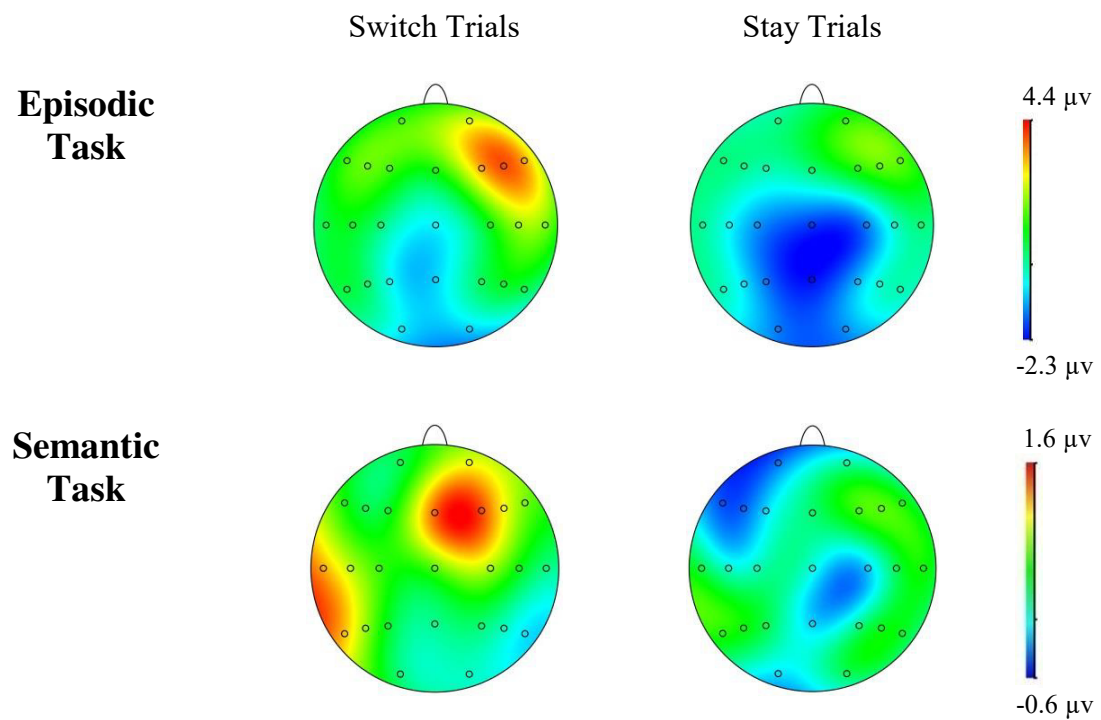
Figure B.4. Topographic maps for Experiment 2 test item data: the scalp distributions of the old/new effects (episodic task) and the repetition effects (semantic task) on switch and stay trials (*A. 500-800ms; B. 900-1200ms*).

10.2.2.1 Episodic task

In the 500-800ms epoch, there was a main effect of old/new status ($F(1, 22) = 13.66$, $p < 0.001$, $d_z = 0.77$, 78% CL) moderated by an old/new by switch/stay interaction ($F(1, 22) = 9.16$, $p < 0.05$, $\eta_p^2 = 0.29$). Separate ANOVAs were conducted for switch and stay trials, and these indicated that the interaction reflects a larger old/new effect on switch trials: there was a reliable main effect of old/new status on switch trials only ($F(1, 22) = 19.90$, $p < 0.001$, $d_z = 0.93$, 82% CL). For stay trials there was only a trend for the main effect of old/new status ($F(1, 22) = 4.03$, $p = 0.057$, $d_z = 0.42$, 66% CL). In the 900-1200ms epoch, there was a main effect of old/new status ($F(1, 22) = 14.03$, $p < 0.001$, $d_z = 0.78$, 78% CL) moderated by interactions between old/new status and hemisphere ($F(1, 22) = 19.18$, $p < 0.001$, $\eta_p^2 = 0.47$) and old/new status and site ($F(1, 22) = 7.19$, $p < 0.05$, $\eta_p^2 = 0.25$). These interactions reflect a positive-going old/new effect which was somewhat larger for the mid-lateral sites than superior sites and over the right than left hemisphere. There was also a trend for an interaction between old/new status and switch/stay ($F(1, 22) = 3.38$, $p = 0.08$, $\eta_p^2 = 0.13$).
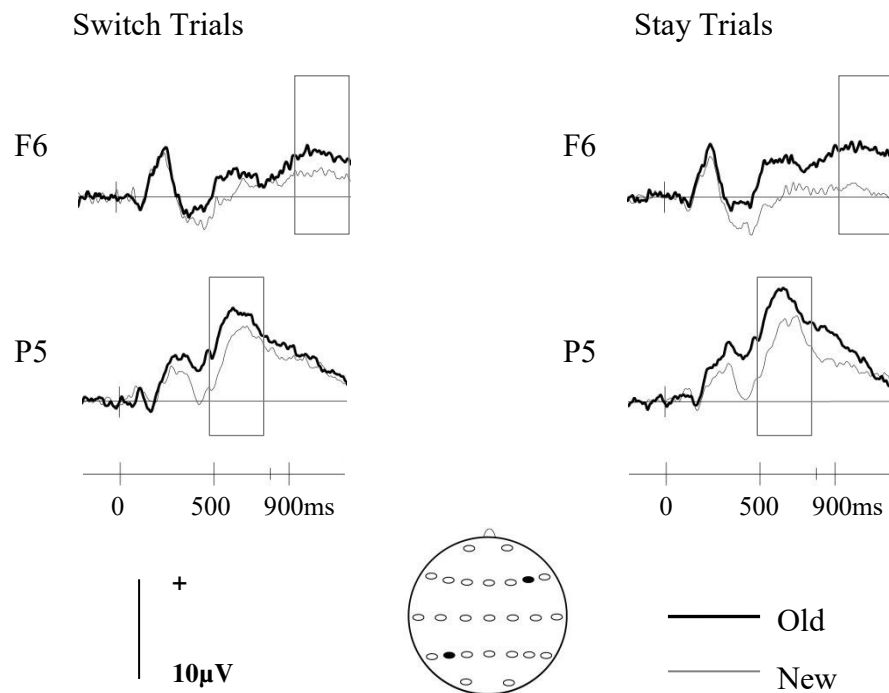
10.2.2.2 Semantic task

In the 500-800ms epoch, there was a main effect of old/new status ($F(1, 22) = 7.61$, $p < 0.05$, $d_z = 0.58$, 72% CL) moderated by an old/new by hemisphere interaction ($F(1, 22) = 5.92$, $p < 0.05$, $\eta_p^2 = 0.21$) reflecting a positive-going old/new effect that did not vary with switch/stay status and was somewhat larger over the left than right hemisphere. In the 900-1200ms epoch, there was an interaction between old/new status and hemisphere ($F(1, 22) = 6.24$, $p < 0.05$, $\eta_p^2 = 0.22$) reflecting greater relative positivity for old than new items over the right hemisphere and greater relative positivity for new items over the left hemisphere. There was also a trend for an interaction between old/new status, switch/stay, hemisphere and site ($F(1, 22) = 3.54$, $p = 0.073$, $\eta_p^2 = 0.14$).

10.2.3 Experiment 3

This analysis included the data from a subset of 23 participants where sufficient trial numbers (≥16) were available for comparisons in each of the conditions of interest. On average, 84% of the available trials contributed to the ERP test item data in this experiment for each participant. For the test item data the mean trial numbers contributing to the ERPs (ranges in parenthesis) were: episodic switch old = 22 (17-28), episodic stay old = 25 (20-30), episodic switch new = 22 (13-27), episodic stay new = 25 (16-28), semantic switch old = 25 (17-30), semantic stay old = 29 (25-30), semantic switch new = 26 (19-30), semantic stay new = 28 (22-30).

Figure B.5 demonstrates the grand averaged ERP waveforms associated with item status (old/new) for each task at a representative right anterior (F6) and left posterior (P5) site, separated for switch and stay trials. Scalp maps depicting the ERP old/new effects (episodic task) and repetition effects (semantic task) are shown for each epoch in **Figure B.6** (A. 500-800ms, B. 900-1200ms) separated for switch and stay trials.
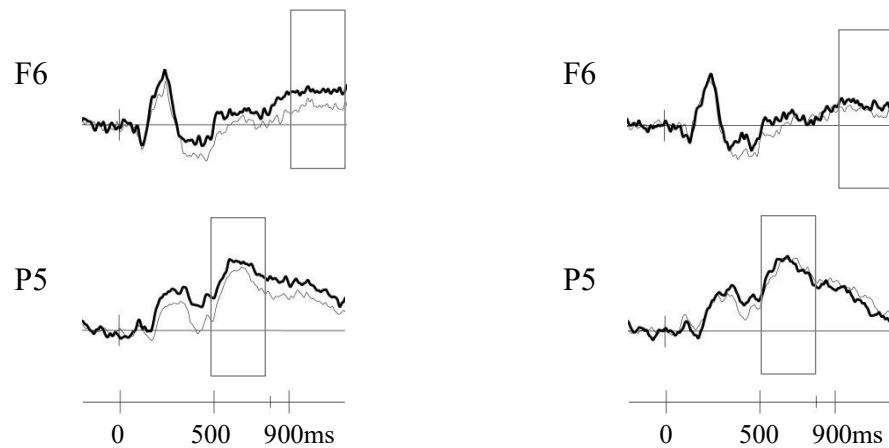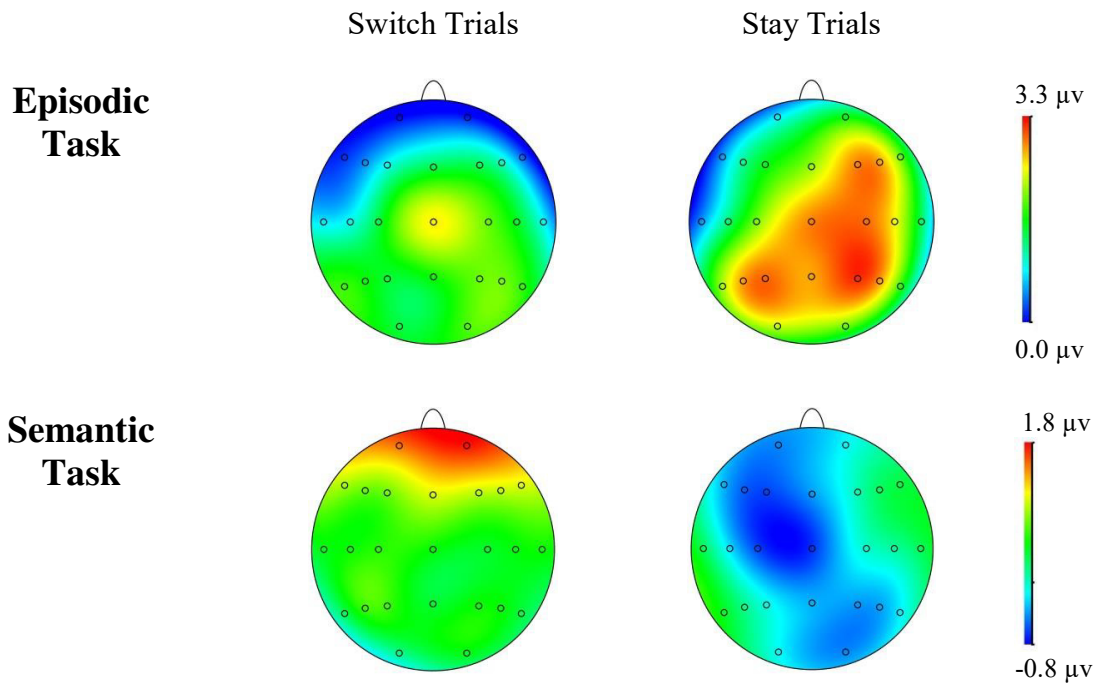
Figure B.5. Grand averaged ERPs associated with old and new test items in Experiment 3 for switch and stay trials at a representative right anterior (F6) and left posterior electrode site (P5). For the semantic task the data are separated only by old/new status. For the episodic task the data are separated for items attracting correct old or new responses.
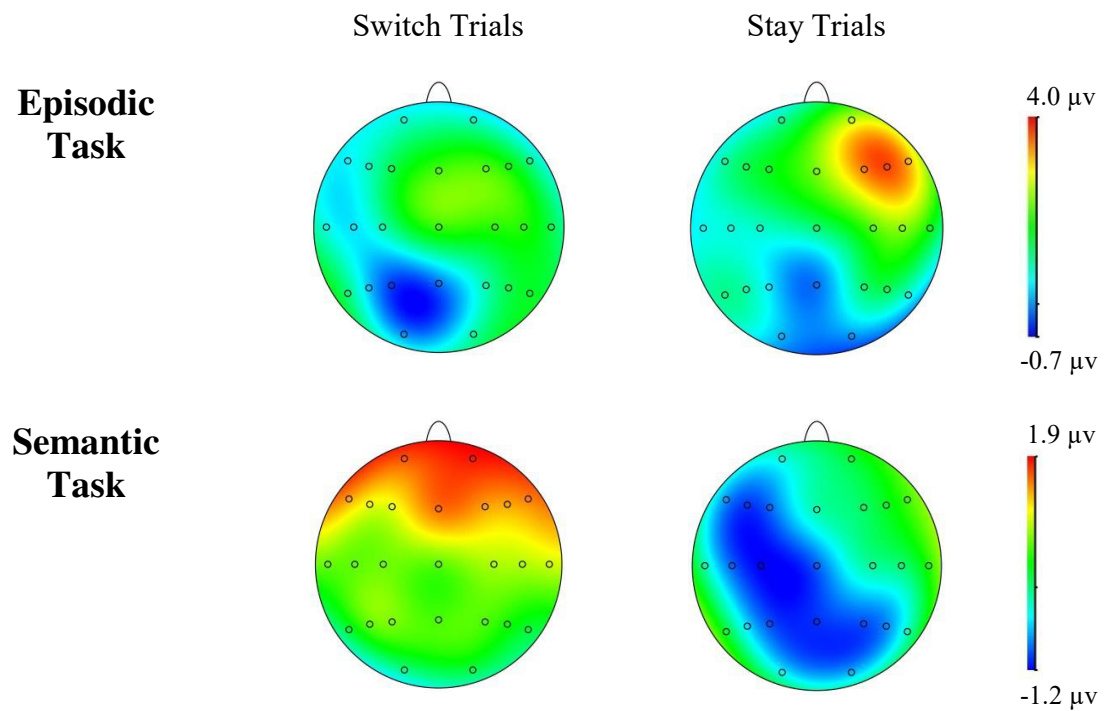
Figure B.6. Topographic maps for Experiment 3 test item data: the scalp distributions of the old/new effects (episodic task) and the repetition effects (semantic task) on switch and stay trials (*A. 500-800ms; B. 900-1200ms*).

10.2.3.1 Episodic task

In the 500-800ms epoch, there was a main effect of old/new status ($F(1, 22) = 20.52$, $p < 0.001$, $d_z = 0.94$, 83% CL), as well as a trend for an interaction between old/new status, switch/stay and site ($F(1, 22) = 3.49$, $p = 0.075$, $\eta_p^2 = 0.14$). In the 900-1200ms epoch, there was a main effect of old/new status ($F(1, 22) = 15.59$, $p < 0.05$, $d_z = 0.82$, 79% CL) moderated by an old/new by hemisphere interaction ($F(1, 22) = 10.24$, $p < 0.05$, $\eta_p^2 = 0.32$) reflecting a positive-going old/new effect which was somewhat larger over the right than left hemisphere. There was also a trend for an interaction between old/new status, switch/stay and hemisphere ($F(1, 22) = 3.04$, $p = 0.096$, $\eta_p^2 = 0.12$).

10.2.3.2 Semantic task

In the 500-800ms epoch, there was only a trend for an interaction between old/new status, hemisphere and site ($F(1, 22) = 2.97$, $p = 0.099$, $\eta_p^2 = 0.12$). In the 900-1200ms epoch, there was an interaction between old/new status and hemisphere ($F(1, 22) = 7.68$, $p < 0.05$, $\eta_p^2 = 0.26$) reflecting a positive-going old/new effect which was evident over the right and not the left hemisphere. There were also trends for interactions between old/new status, switch/stay, hemisphere and site ($F(1, 22) = 3.44$, $p = 0.078$, $\eta_p^2 = 0.14$) and old/new status and switch/stay ($F(1, 22) = 3.02$, $p = 0.096$, $\eta_p^2 = 0.12$).

10.2.4 Experiment 4

This analysis included the data from a subset of 10 participants where sufficient trial numbers (≥16) were available for comparisons in each of the conditions of interest. On average, 89% of the available trials contributed to the ERP test item data in this experiment for each participant. For the test item data the mean trial numbers contributing to the ERPs (ranges in parenthesis) were: episodic switch old = 26 (23-29), episodic stay old = 18 (16-20), episodic switch new = 22 (17-28), episodic stay new = 18 (16-20), semantic switch old = 28 (26-29), semantic stay old = 19 (18-20), semantic switch new = 27 (26-29), semantic stay new = 19 (17-20).

Figure B.7 demonstrates the grand averaged ERP waveforms associated with item status (old/new) for each task at a representative right anterior (F6) and left posterior (P5) site, separated for switch and stay trials. Scalp maps depicting the ERP old/new effects (episodic task) and repetition effects (semantic task) are shown for each epoch in **Figure B.8** (A. 500-800ms, B. 900-1200ms) separated for switch and stay trials.
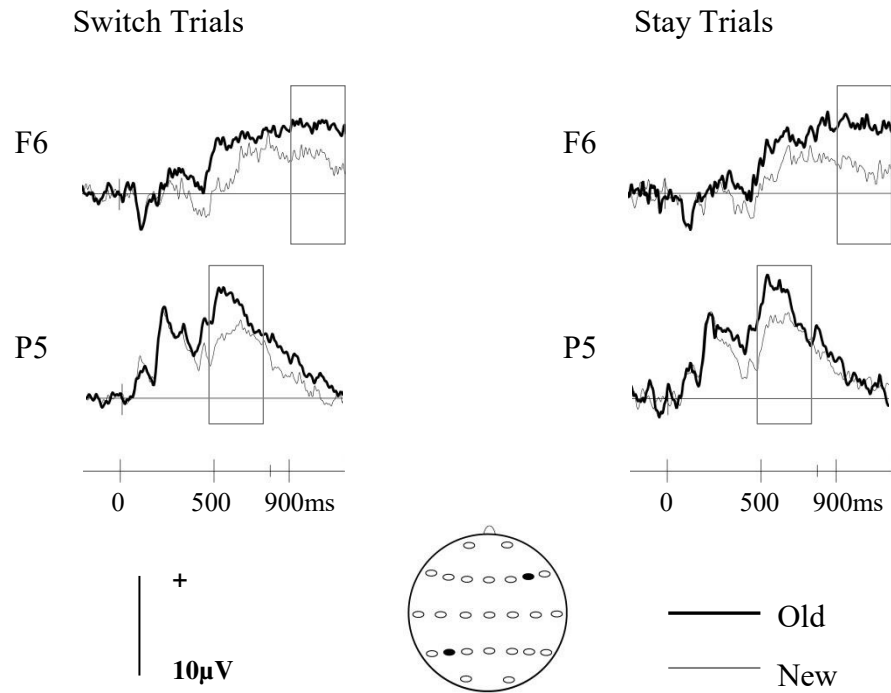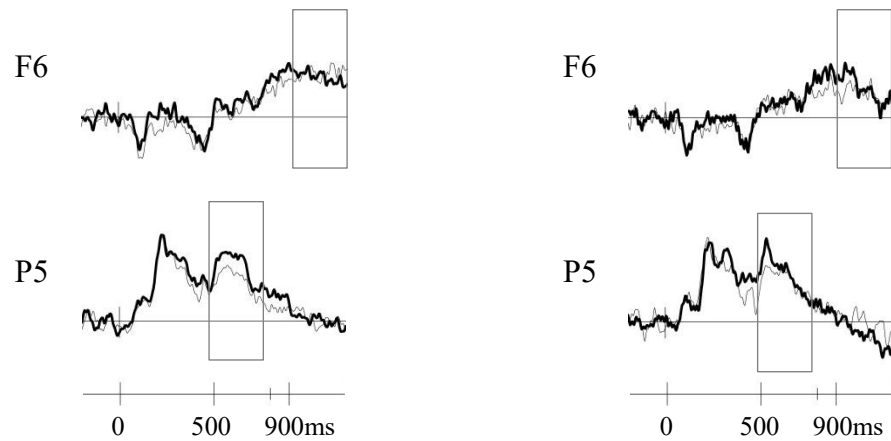
Figure B.7. Grand averaged ERPs associated with old and new test items in Experiment 4 for switch and stay trials at a representative right anterior (F6) and left posterior electrode site (P5). For the semantic task the data are separated only by old/new status. For the episodic task the data are separated for items attracting correct old or new responses.
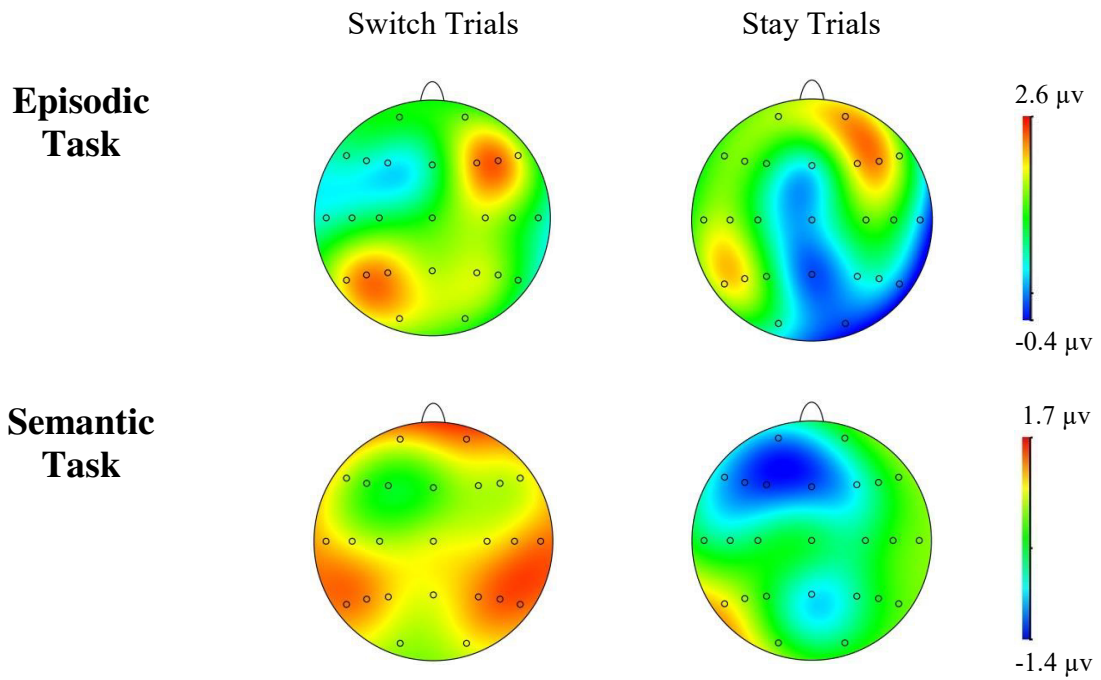
Figure B.8. Topographic maps for Experiment 4 test item data: the scalp distributions of the old/new effects (episodic task) and the repetition effects (semantic task) on switch and stay trials (*A. 500-800ms; B. 900-1200ms*).

10.2.4.1 Episodic task

In the 500-800ms epoch, there was an interaction between old/new status and hemisphere ($F(1, 9) = 5.19$, $p < 0.05$, $\eta_p^2 = 0.37$; and a trend for a main effect of old/new status: $F(1, 9) = 3.50$, $p = 0.094$, $d_z = 0.39$, 65% CL) reflecting a positive-going old/new effe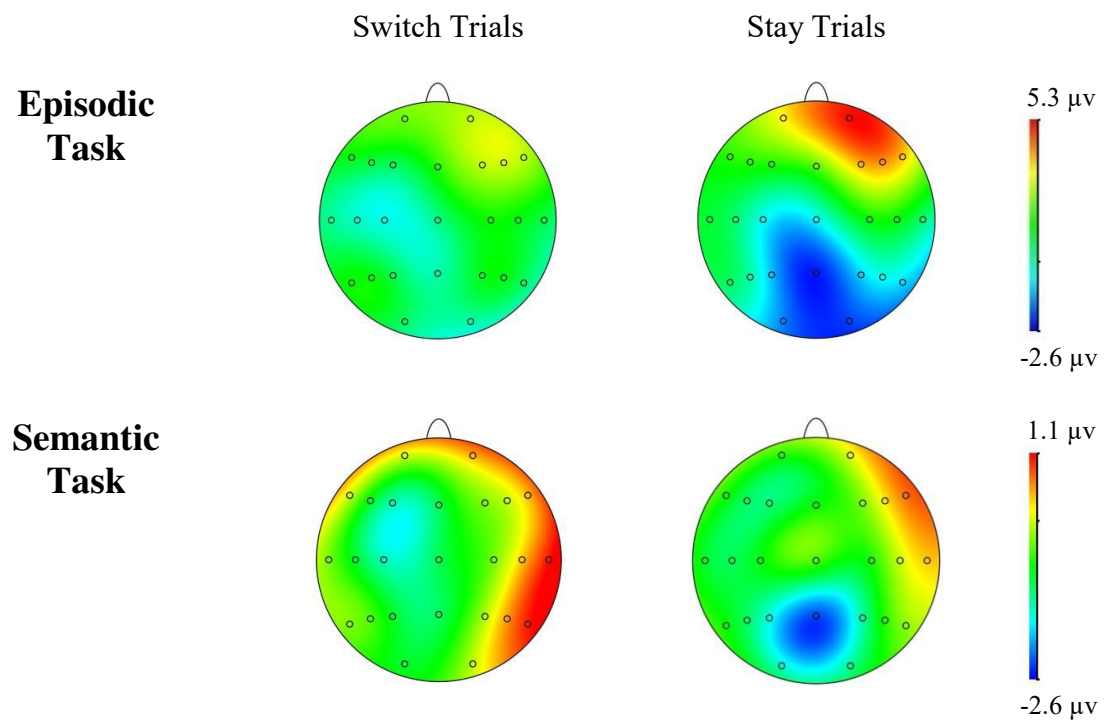ct that did not vary with switch/stay status and was somewhat larger over the left than right hemisphere. In the 900-1200ms epoch, there were interactions between old/new status and hemisphere ($F(1, 9) = 6.43$, $p < 0.05$, $\eta_p^2 = 0.42$) and old/new status, switch/stay, and site ($F(1, 9) = 5.24$, $p < 0.05$, $\eta_p^2 = 0.37$) (and a trend for a main effect of old/new status: $F(1, 9) = 4.64$, $p = 0.06$, $d_z = 0.45$, 67% CL). Separate ANOVAs were conducted for switch and stay trials, and these indicated that the interaction reflects a larger old/new effect on stay trials: there was a reliable main effect of old/new status on stay trials only ($F(1, 9) = 9.21$, $p < 0.05$, $d_z = 0.96$, 83% CL).

10.2.4.2 Semantic task

In the 500-800ms epoch, there were no reliable old/new effects. In the 900-1200ms epoch, there was an interaction between old/new status and hemisphere ($F(1, 9) = 5.66$, $p < 0.05$, $\eta_p^2 = 0.39$). This reflects a greater relative positivity for new than old items over both the left and right hemisphere, which was somewhat larger over the left than right.

10.2.5 Summary

For the 500-800ms epoch, it was anticipated that old/new effects would be larger on stay than switch trials for the episodic task (Evans et al., 2012; Wilckens et al., 2011), and repetition effects would be larger on switch than stay trials for the semantic task (Evans, Herron, et al., 2015). This kind of crossover was not evident in any of the four experiments individually, and overall the statistical evidence for changes in the magnitudes of effects in this epoch was limited. In Experiment 1 there was an indication that the repetition effect was larger on stay than on switch trials, while in Experiment 2 there was some evidence that the old/new effect was larger on switch than on stay trials.

No strong *a priori* predictions were made for the analyses of right-frontal effects in a 900-1200ms time window, and as for the analyses in the earlier epoch no consistent pattern of reliable outcomes involving the switch/stay dimension were obtained. There was statistical evidence for a larger old/new effect on stay than on switch trials in Experiment 4 only.

It may well be the case that differences between the task demands in the four experiments would lead to different engagement of processes contributing to the quality/quantity of information retrieved from memory, as well as to operations that are active at a later point in the processing pipeline. The statistical outcomes in these experiments, however, do not offer much insight into this, in all likelihood due to the somewhat low participant and average trial numbers contributing to the conditions of interest. The fact that trial number constraints also precluded separating the data according to the accuracy of source judgments is also notable: both left-parietal and right-frontal old/new effects have been shown to be sensitive to source accuracy (Smith, 1993; Wilding, Doyle, & Rugg, 1995; Wilding & Rugg, 1996).

Additional Bayesian replication tests of the earlier Herron and Wilding (2004, 2006a) effects were conducted for the data from Evans, Williams, et al. (2015). Here, the divergence from Evans, Williams, et al. (2015) was obtained on switch trials, and was input as a replication of the 'location' effects identified on stay trials in the earlier studies. For Evans, Williams, et al. (2015) the *t* value was calculated using the averages for each cue-type across the three right-frontal sites, for switch trials (800-1900ms). There was one episodic location task, and one perceptual judgment task. The number of participants and the *t* value for the main effect of interest, in the original and replication experiments are shown in **Table C.1**.

Table C.1. *t* values and sample sizes (*N*) from the original studies demonstrating a greater right-frontal positivity during preparation for episodic than for semantic memory retrieval (Herron and Wilding, 2004; 2006a), and for the replication attempt (Evans, Williams, et al., 2015).

Study	*t* value	*N*
Herron and Wilding (2004): Main effect of cue-type (location/semantic)	2.86	20
Herron and Wilding (2006a): Main effect of cue-type (location/semantic)	2.47	16
Evans, Williams, et al. (2015): Main effect of cue-type (location/perceptual)	3.09	32

For Evans, Williams, et al. (2015) as a replication of the episodic retrieval mode location effect in Herron and Wilding (2004) the BF = 40.42 (**Figure C.1A**). For a replication of the location effect in Herron and Wilding (2006a) the BF = 38.51 (**Figure C.1B**). Thus, the effect obtained on switch trials in Evans, Williams, et al. (2015) provides very strong evidence for the alternative hypothesis of the effects obtained in the earlier Herron and Wilding (2004, 2006a) papers.

A. Evans, Williams, et al. (2015): for 2004 'location' effect

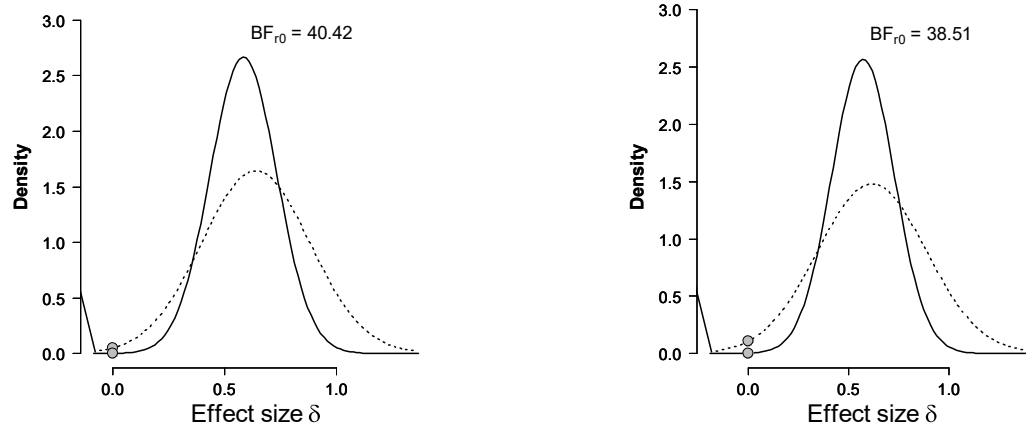**B.** Evans, Williams, et al. (2015): for 2006a 'location effect'

Figure C.1. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the right-frontal positivity identified previously during preparation for episodic memory retrieval. In each panel, the dotted lines represent the posterior from the original study (A. Herron & Wilding, 2004; B. Herron & Wilding, 2006a), which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (Evans, Williams, et al., 2015) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

Bayesian replication tests (Verhagen & Wagenmakers, 2014) were then also carried out in order to investigate whether the ERP results from Experiments 1-4 provide support for the null or alternative hypothesis: for more positive-going right-frontal ERP activity on switch trials, when preparing for episodic in comparison to non-episodic tasks. The number of participants and the *t* value for the main effect of interest, in the original and replication experiments are shown in **Table C.2**.

Table C.2. *t* values and sample sizes (*N*) from the original study demonstrating right-frontal positivity during preparation for episodic memory retrieval on switch trials (Evans, Williams, et al., 2015), and for the replication attempts (Experiments 1-4).

Study	*t* value	*N*
Evans, Williams, et al. (2015):		32
Main effect of cue-type (location/perceptual)	3.09	
Experiment 1:		24
Analysis strategy as 2015 study	-1.03	
Experiment 2:		32
Analysis strategy as 2015 study	0.88	
Experiment 3:		32
Analysis strategy as 2015 study	-0.80	
Experiment 4:		32
Analysis strategy as 2015 study	1.51	

For Experiment 1, as a replication of the divergence obtained on switch trials in Evans, Williams, et al. (2015) the BF = 0.03 (**Figure C.2A**), providing very strong evidence for the null hypothesis. For Experiment 2, as a replication of the divergence identified on switch trials in Evans, Williams, et al. (2015) the BF = 0.33 (**Figure C.2B**), providing substantial evidence for the null hypothesis. For Experiment 3, as a replication of the divergence identified on switch trials in Evans, Williams, et al. (2015) the BF = 0.03 (**Figure C.2C**), providing very strong evidence for the null hypothesis.

However, for Experiment 4, as a replication of the divergence identified on switch trials in Evans, Williams, et al. (2015) the BF = 1.2, (**Figure C.2D**) providing anecdotal evidence for the alternative hypothesis. Importantly, with 24 participants this comparison indicated anecdotal evidence for the alternative hypothesis, thus, in line with the Bayesian Stopping Rule (Dienes, 2011; Verhagen & Wagenmakers, 2014) data collection was increased to 32 participants. However, the results remained the same (anecdotal evidence for the alternative hypothesis) and data collection was terminated at this point.
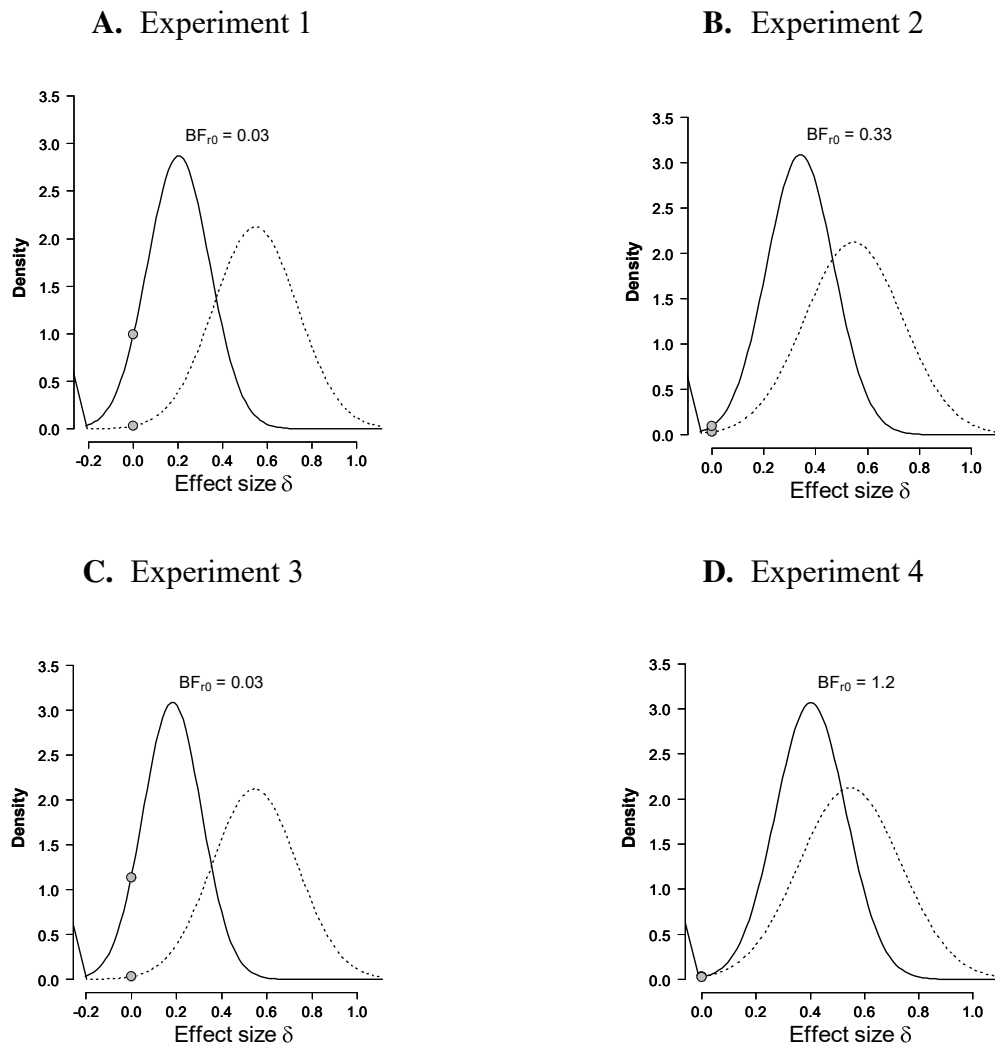
A. Experiment 1

B. Experiment 2

C. Experiment 3

D. Experiment 4

Figure C.2. Bayesian results of the Replication Test (Verhagen & Wagenmakers, 2014) for the right-frontal positivity identified previously during preparation for episodic memory retrieval on switch trials. In each panel, the dotted lines represent the posterior from the original study (Evans, Williams, et al., 2015), which was used as the prior for the effect sizes in the replication tests. The solid lines represent the posterior distributions after the data from the replication attempt (A-D. Experiments 1-4) are taken into account. The grey dots indicate the ordinates of this prior and posterior for the null hypothesis that the effect size is zero. The ratio of these two ordinates gives the result of the replication test (Verhagen & Wagenmakers, 2014).

10.4 Appendix D. Behavioural results summary table: Experiments 1-6

Table D.1. Behavioural results for Experiments 1-6.

	Experiment 1: matched		Experiment 2: unmatched		Experiment 3: words & unmatched		Experiment 4: 500ms RCI & unpredictable		Experiment 5: stay+1			Experiment 6: alternating & blocked		
Encoding Accuracy	0.96		0.98		0.98		0.97		0.96			0.98		
	Switch	Stay	Switch	Stay	Switch	Stay	Switch	Stay	Switch	Stay	Stay+1	Switch	Stay	Blocked
Accuracy														
Episodic														
Old/new discrimination (*Pr*)	0.58	0.65	0.61	0.68	0.65	0.69	0.68	0.72	0.62	0.69	0.66	0.61	0.70	0.71
P(correct source)	0.75	0.81	0.76	0.77	0.73	0.75	0.76	0.83	0.73	0.74	0.76	0.82	0.80	0.83
Hit	0.85	0.87	0.86	0.85	0.83	0.85	0.89	0.88	0.85	0.85	0.83	0.81	0.84	0.83
Correct rejection	0.73	0.78	0.75	0.83	0.83	0.84	0.79	0.84	0.77	0.83	0.83	0.80	0.86	0.88
Semantic														
Correct classification	0.73	0.73	0.75	0.75	0.79	0.76	0.75	0.74	0.67	0.67	0.70	0.70	0.71	0.70
Reaction Time (ms)														
Episodic														
Correct source	1509	1412	1540	1539	1656	1551	1521	1441	1542	1505	1447	1462	1420	1259
Correct new	1327	1226	1351	1227	1332	1215	1312	1236	1318	1241	1180	1265	1160	977
Semantic														
Correct old	1483	1342	1375	1339	1395	1413	1573	1480	1537	1524	1451	1545	1405	1192
Correct new	1385	1367	1293	1271	1359	1305	1500	1389	1529	1469	1454	1363	1344	1206