Recognition-induced updating of face memories: behavioural and electrophysiological evidence

Matthew Luke Plummer

School of Psychology

University of Kent

A thesis submitted for the Degree of Ph.D. in Cognitive Psychology/Neuropsychology in the Faculty of Social Sciences at the University of Kent

July 2020

Word count: 73,971

Abstract

Episodic memories are prone to ‘updating’, that is, memories can be strengthened or distorted after their initial encoding. The majority of research has examined the cognitive and neurocognitive mechanisms for the updating of elaborate episodic memories, however the goal of this thesis was to examine mechanisms for the updating of simpler episodic memories, such as memories for faces. For all experiments, a novel repeated recognition paradigm required participants to complete two recognition tests, with target faces (shown during a previous learning phase) presented amongst four distractor faces (not seen prior to the first test). Critically, face stimuli were derived from artificial face space models to control perceptual differences between images, as well as to use the Euclidean distance between face images as a continuous metric of recognition (details in Chapter 2). Within Chapter 3, it was found that elevated confidence judgements during initial recognition attempts predicted whether participants would recognise the same face in a subsequent test, regardless of the accuracy of recognition judgements on the initial test. In Chapter 4, it was queried whether face memory updating would be increased after retrieving vs. re-studying face memories. Results showed that retrieval enhanced the updating of face memories compared to re-study tasks, despite these tasks encouraging participants to encode faces that were cued to participants (Experiment 4a & c) or were selected according to distinctiveness (Experiment 4c). Finally, the electrophysiological correlates of face memory retrieval and updating were examined with ERPs (Chapter 5) and oscillations (Chapter 6). ERPs largely corresponded to the retrieval and reactivation of target memories recognised with high confidence. However, oscillatory markers of objective, subjective and updating processes were found. Together, this thesis provides the first evidence of the cognitive and neurocognitive mechanisms underlying the retrieval-induced updating of face memories.

Acknowledgements

I first want to thank Jaimee for her constant love, support and encouragement during my PhD, as well as to my family for backing me from day one (especially Mum, Dad & Alison, Gran & Grandad, John & Andy, Vic, Andy, AK & Niamh). I also want to thank my friends who I have shared this journey with, including but not limited to Louisa, Mel and Rebecca, to Robin, Louisa and Fadi, and to Hannah and Nat.

I want to thank Dr Zara Bergström for her excellent supervision, support and guidance throughout the past four years, in addition to all of the people who I have met and worked with for helping me along the way, you know who you are!

Declaration

I declare that this thesis is my own work carried out under the normal terms of supervision.

…………………………

Matthew Luke Plummer

Contents

Chapter 1 - General introduction and literature review

Episodic memory retrieval allows us to relive personal experiences from our past (Tulving, 1972). The process of retrieval is thought to involve active, reconstructive mechanisms that can modify and *update* the retrieved memory (Dudai & Eisenberg, 2004). There are two consequences that retrieval has on the updating of episodic memories. When a memory is correctly retrieved, the likelihood of this memory being remembered correctly during subsequent retrieval attempts is increased. In contrast, if memories are not remembered accurately (for example, certain details of a memory are forgotten, or incorrect details are retrieved), then this modified, distorted memory will be continuously retrieved again at the expense of retrieving the original episode (Hardt, Einarsson & Nader, 2010). In the context of this thesis episodic memory updating is defined as the selection of information during retrieval attempts that is the same/similar to information that is selected during previous retrieval or re-study tasks.

In order to understand why memories become updated, research has examined the cognitive and neurocognitive mechanisms underlying this phenomenon. For example, a series of studies from Bridge and Paller (2012) and Bridge and Voss (2014) examined the neurocognitive mechanisms underlying the updating of objects paired with a spatial context, using cued-recall paradigms. Furthermore, mechanisms of updating have also been studied using autobiographical memories containing rich, vivid details of a memory (St. Jacques, Olm & Schacter, 2013; St. Jacques & Schacter, 2013). Whilst these studies have been pivotal to understand the mechanisms of memory updating, much less research has examined the updating of simpler forms of episodic memory, such as memories for items that contain minimal semantic content and have few contextual associations. To address this issue, this thesis will focus on studying the updating of face memories. In the context of the present thesis, face memories lack semantic/conceptual information in the sense that we do not know any additional information (e.g., name, occupation) for a face that has never been seen before. Furthermore, the face memories in this thesis are considered to have minimal contextual associations given that participants would have only ever seen faces during learning or recognition tasks. Thus, the only contextual information associated with target faces is the learning context, meaning that the mechanisms of face memory updating may differ to the updating of memories with rich contextual associations.

Previous research has demonstrated how repeated retrieval attempts can update eyewitness memories of criminal suspects who were previously unknown to the witness (Steblay & Dysart, 2013). However, less research has examined the cognitive or neurocognitive factors that contribute to face memory updating. This is an important issue to address as the mechanisms underlying the recognition-induced updating of face memories may differ to the mechanisms of elaborate episodic memory updating, given the differences in brain regions thought to be related to these different types of memory (Ranganath & Ritchey, 2012), as well as differences in neural activity based on the semantic content of the retrieved memory (e.g. Mackenzie, Alexandrou, Hancock & Donaldson, 2018; Mackenzie & Donaldson, 2007; 2009, Nie, Griffin, Keinath, Walsh, Dittmann & Reder, 2014). Consequently, it is important for the research in this thesis to examine neurocognitive factors that underlie the updating of face memories to advance our understanding of this phenomenon.

In addressing the central aim of this thesis, evidence of face memory updating will be presented across five experimental chapters. In three of these chapters, it was investigated whether subjective retrieval processes modulate the updating of face memories. In the literature, the majority of studies focus on memory updating in relation to the objectiveness of retrieval decisions (i.e. updating of memories following correct or incorrect retrieval judgements). Nevertheless, retrieval attempts are not categorised just by their objectiveness, but also in terms of the subjective experience individuals have during retrieval attempts. Critically, objective and subjective retrieval processes are not concomitant, that is, increased subjective experience during retrieval can occur both when retrieval is accurate and erroneous. Furthermore, studies have shown that updating is more likely for memories that are retrieved with enhanced subjectivity (St. Jacques et al., 2013; St. Jacques & Schacter, 2013). Taking these ideas into this thesis, several experiments were conducted to investigate whether confidence judgements during an initial recognition attempt predicted whether participants would select the same face in a subsequent recognition test. This was investigated both when initial recognition was correct and incorrect, contributing to the investigation of both positive and negative consequences of face memory updating (see Chapter 3). Furthermore, this issue was investigated with the use of electroencephalography (EEG) to examine the electrophysiological correlates of face memory updating by analysing both event-related potentials (ERPs; see Chapter 5) and oscillatory power (see Chapter 6). The aim of these chapters was to develop on the behavioural findings presented in Chapter 3 by examining the neurocognitive mechanisms underlying the updating of face memories.

The second line of research in this thesis questions whether retrieval is a necessary factor to induce face memory updating. A body of research called the ‘retrieval-practice’ literature examines whether memory updating is more likely following tasks where memories are actively retrieved vs. tasks where memories are passively re-studied (without retrieval attempts). It is generally thought that retrieval enhances memory updating compared to re-study (Roediger & Butler, 2011; Roediger & Karpicke, 2006). However, several experiments in this thesis were conducted to verify whether face memory updating was enhanced following retrieval of memories, or was just as likely to occur after faces were re-studied (see Chapter 4). Overall, the factors examined in this thesis are the first, to my current knowledge, that attempt to understand the cognitive and neural mechanisms of the retrieval-induced updating of face memories. In the next section, the thesis will present a literature review of all of the aforementioned factors relevant for this thesis.

For the present thesis, face memory updating was examined following learning procedures designed to maximise the encoding of target faces so that participants would be able to remember and recognise these faces in subsequent recognition tasks. For this, the same target faces were presented multiple times in distinct learning cycles, designed to strengthen face memory representations (either by presenting the same face several times, or by creating multiple contextual associations for each face). However, this methodology questions whether the face memories in the current experiments are episodic in nature (as episodic memories are defined as remembering single events according to Tulving, 1972). Alternatively, the face memories in this thesis may be better considered as 'repeated events', this being the repetition of the same information across multiple episodes (Rubin & Umanath, 2015). Therefore, participants could either recall specific episodes associated with faces (e.g. recollect a specific learning cycle associated with a target face). Alternatively, participants could assess the general gist of having seen a face before (regardless of, and without recalling, specific details from single learning cycles), merging repeated events as one single event (Rubin & Umanath, 2015). Both of these processes require participants to have a mnemonic experience during face recognition, thus could allow for the updating of these memories by several factors such as subjective retrieval processes, or via processes associated with retrieval vs. re-study.

1.1. Episodic memory updating: phenomenological evidence

The updating of episodic memories, both to a positive and negative consequence, has been demonstrated in research examining the retrieval-practice effect, a phenomenon suggesting that the act of retrieval is a powerful mechanism for the learning and updating of retrieved memories. The paradigm used in retrieval practice studies typically presents participants with information in a study phase (e.g. word pairs, such as apple - beach, book - restaurant), after which they complete an intermediate task with one of three conditions; a retrieval condition (remember information from study, e.g. recall word pairs when presented with apple - ____ or book - ____), a re-study condition (re-view information from study, e.g. apple - beach, book - restaurant), or a control condition (information not presented). A final ‘criterial’ test is completed afterwards (e.g. recall word pairs when presented with apple - ____ or book - ____) for items in each previous condition. The key finding from the retrieval-practice paradigm shows that participants are more likely to repeatedly retrieve the same information between intermediate and criterial tasks if the intermediate task required retrieval, compared to re-study or control conditions (Roediger & Butler, 2011). The retrieval practice effect has been shown to be incredibly robust (Roediger & Karpicke, 2006), with effects demonstrated for word lists, paired associates, prose texts, lecture content and non-verbal materials, emphasising the critical nature of retrieval practice in educational contexts (Karpicke & Blunt, 2011; Roediger & Butler, 2011). Furthermore, the retrieval practice effect has been demonstrated with a variety of tasks requiring different memory retrieval demands, including free recall, cued recall and recognition (Roediger & Marsh, 2005).

Of particular relevance to the present thesis, retrieval practice effects have also been found when recognition memory is tested in multiple choice question (MCQ) format (Butler, Marsh, Goode & Roediger, 2006; Marsh, Roediger, Bjork & Bjork, 2007;

Roediger & Marsh, 2005). In these studies, participants typically learn semantic information in the form of educational materials (e.g. “general knowledge facts”). During subsequent MCQ testing, participants are asked questions related to information shown during previous study phases, with multiple response options given. Response options typically consist of the correct answer (seen during study) shown amongst multiple incorrect answers (not seen during study). Participants are required to select one of these options that is consistent with what they remember from the study phase. A re-study task using a similar format can be created by presenting multiple response options and telling participants to select one of these options to be remembered for subsequent memory tests. A criterial test then follows, allowing the examination of how retrieval/re-study performance during prior MCQ formats influences subsequent memory performance.

Interestingly, the impact of MCQ testing on criterial retrieval performance has opposing consequences. Positive MCQ testing effects show improved retrieval performance during criterial retrieval when participants accurately recognised the correct answer during initial MCQ’s, compared to previously re-studying the correct answer. On the other hand, negative MCQ testing effects show that participants are more likely to make a recognition error during criterial testing when participants falsely selected an incorrect answer during the previous MCQ, compared to not having an MCQ previously (Marsh et al., 2007; Roediger & Marsh, 2005). These results emphasise the beneficial and detrimental impact of initial recognition on the updating of memories containing rich semantic/conceptual information.

Whilst the retrieval practice effect shows the malleable nature of semantically rich, complex episodic memories, this research has not addressed whether similar updating occurs of simpler memories containing minimal semantic/conceptual information, for example, memories for faces. This aspect of memory updating has been addressed in research examining eyewitnesses' memories of a person's face that they saw commit a crime. Eyewitnesses may be asked to remember a suspect's face using different methods, including creating a composite image of a suspect's face, or identifying a suspect shown amongst several innocent individuals through line-ups, photo-arrays, mugbooks or show-ups. According to Steblay and Dysart (2016), it is not uncommon for an eyewitness to be asked to complete several suspect identifications, by either repeating the same identification procedure, or being asked to complete different identification procedures throughout police investigations. Asking eyewitnesses to complete several identification attempts may be an intuitive method of assessing eyewitness's memory of a crime. Indeed, Steblay and Desart (2016) argue that if an eyewitness identifies the same individual across multiple face identifications, then police may consider the eyewitnesses memory of the crime to be reliable. However, in recent decades, the reliability of eyewitness memory has come under scrutiny. On one hand, Wixted, Mickes and Fisher (2018) argue that an eyewitnesses memory can be trusted so long as eyewitnesses are confident in their initial identifications, and that the police follow optimal procedures by providing 'fair' line-ups such as ensuring that one face does not possess any distinctive perceptual features from the rest of the faces in the line-up (Yates, 2017). Nevertheless, the reliability of eyewitness memory has come into question with evidence showing that an innocent person may be wrongly identified across multiple identification procedures (Steblay & Dysart, 2016), which may partly be due to updating of face memories as a result of repeated retrieval.

The consequences of repeated identification procedures, both positive and negative, have been tested in laboratory scenarios to understand factors that predict repeated identifications across multiple identification attempts. From this body of work, studies have defined the repeated selection of a suspect as the "commitment effect". It is important to emphasise that commitment effects applies to both repeatedly identifying the guilty suspect, as well as repeatedly identifying an innocent suspect. Both types of commitment effect have been demonstrated using various combinations of face identification procedures, including the use of mugbooks prior to line-ups (Deffenbacher, Bornstein & Penrod, 2006; Goodsell, Neuschatz & Gronlund, 2009), show-ups prior to line-ups (Haw, Dickinson & Meissner, 2007; Lawson & Dysart, 2012; Valentine, Davis, Memon & Roberts, 2011), photo array's prior to line-ups (Pezdek & Blandon-Gitlin, 2005) and composite construction prior to line-ups (Davis, Gibson & Solomon, 2014; Davis, Maigut, Jolliffe, Gibson & Solomon, 2015; Kempen & Tredoux, 2012; Topp-Manriquez, McQuinston & Malpass, 2016). Several theoretical perspectives have been proposed to account the commitment effect for innocent suspects (see Steblay & Dysart, 2016), including source confusion (witnesses remember seeing an innocent face but misattribute to the crime context), demand characteristics (wanting to commit to the same selection to satisfy police investigations) or through suggestive means (witnesses select faces that are consistent to both line-ups). However, these studies have not provided evidence that can distinguish between the different theoretical accounts as to why an eyewitness memory for a suspect's face seems to become modified through repeated identification procedures.

Furthermore, the methodological nature of these laboratory-based eyewitness experiments, whilst maintaining a high ecological validity to mimic real-world criminal procedures, may introduce confounds that influence the interpretation of face memory updating. For example, most applied experiments will ask participants to passively view a mock crime, thus there is no guarantee that participants are always encoding the suspect's face. Therefore, evidence from these studies cannot be used to conclude that stored face memories have been updated by retrieval, since commitment effects across repeated identification tasks might occur only when the original face was not encoded to begin with. Furthermore, during identification tasks (including line-ups, mugbooks or show-ups), there is little control over how many of the presented faces participants attend to, nor how long participants view each face for (except for Pezdek & Blandon-Gitlin, 2005). That is, witnesses may only focus and attend to face images up until they recognise a face they believe belongs to the suspect, therefore any faces presented after the recognised face would not be attended too. Furthermore, witnesses may only glance at faces they do not recognise, yet other face images may be attended to for longer if the face evokes a sense of familiarity to the witness. Therefore, the extent to which faces (presented during intermediate identification tasks) are encoded into long-term memory is varied. Consequently, despite high levels of ecological validity, such factors complicate the theoretical interpretation of results from the applied commitment effect literature. Therefore, the central aim of this thesis was to better understand the neurocognitive mechanisms underlying updating of face memories by adopting methodological approaches from experimental cognitive psychology and cognitive neuroscience research into episodic memory.

1.2 Episodic memory: encoding, retrieval, and updating

To understand the mechanisms underlying the retrieval practice and commitment effects described above, models of episodic memory need to be considered. The traditional view of episodic memory processing suggests that encoding (i.e. the formation of episodes) and retrieval (i.e. remembering prior events) are two distinct processes (Rugg, Johnson & Uncapher, 2015). However, these two processes are thought to overlap, that is, retrieval processes could be engaged during encoding to facilitate learning and storage of memories. Alternatively, encoding may be engaged during retrieval as a possible mechanism of memory updating. For example, applied to the commitment effects described previously, if an eyewitness is presented with a line-up containing several faces, they will be attempting to retrieve their memory of the suspect’s face, whilst at the same time encoding the faces of individuals presented within the line-up. However, there are several factors that determine whether information is encoded during retrieval attempts and thereby leads to memory updating.

Some theoretical accounts posit that the ‘*reactivation’* of a memory is critical for the memory to become updated. Specifically, successful memory retrieval is considered to involve reactivation of a dormant neural representation of a memory in most models of memory retrieval. For complex, episodic memories, evidence suggests that memory reactivation involves reinstatement of activity in regions of the brain that were active during initial encoding of episodic memories (Danker & Anderson, 2010). According to neurocomputational models of memory, such as the Complementary Learning Systems (CLS) models (McClelland, McNaughton & O’Reilly, 1995; Norman & O’Reilly, 2003; O’Reilly, Bhattacharyya, Howard & Ketz, 2011), it is thought that reactivation of memories is initiated in the hippocampus (Ranganath, 2010), a region that stores unique representations for memories. The hippocampus is thought to initiate reinstatement of neural activity in brain regions active during the encoding of memories via a process of pattern completion (Rolls, 2016; see Johnson & Rugg, 2007; Lee, Samide, Richter & Kuhl, 2018; Schapiro et al., 2018; Staresina, Henson, Kriegeskorte & Alink, 2012; Staresina, Alink, Kriegeskorte & Henson, 2013; Van den Honert, McCarthy & Johnson, 2016, for fMRI evidence of memory reinstatement during retrieval).

The idea that reactivated memories are subject to updating has been supported by a series of experiments by Hupbach and colleagues. Firstly, Hupbach, Gomez, Hardt and Nadel (2007) asked participants to learn a list of words (list 1). On a following day, participants then learnt a second list of words in a ‘reminder’ or ‘no reminder’ condition. In the reminder condition, participants were reminded of the task they completed during list 1 encoding (without explicitly asking participants to recall words from list 1) whereas those in the no reminder condition did not receive a list 1 reminder. A free recall task for list 1 and list 2 was then completed immediately or a day following the learning of list 2 words. Interestingly, free recall performance on the following day only showed an asymmetric pattern of ‘memory intrusions’, such that participants were more likely to misattribute list 2 words to list 1 in the reminder versus no reminder condition. However, list 1 items were no more likely to be misattributed to list 2 between reminder and no reminder conditions. The results therefore suggest that reactivation of the list 1 memory in the reminder condition led to the modification of this memory by items from list 2. In sum, this line of research (see Scully, Napper & Hupbach, 2017; Schwabe, Nader & Pruessner, 2014 for reviews) emphasise that reactivation of episodic memories leads to the updating of these memories by information presented during the time-limited reactivation period.

Theoretically, the updating effects observed by Hupbach and colleagues (Capelo, Albuquerque & Cadavid, 2019; Hupbach et al., 2007; Hupbach, Hardt, Gomez & Nadel, 2008) have been attributed to process of reconsolidation (Dudai & Eisenberg, 2004; Hardt, Einarsson & Nadel, 2010). Reconsolidation theory suggests that the reactivation of memories (induced by retrieval or reminder of an event) transforms permanently stored memories into an ‘active’ labile and unstable state for a short period of time. During this time-limited instability, memory traces can be modified by information, including but not limited to mnemonic information generated during retrieval, or information present in the retrieval environment (such as list 2 items in the studies by Hupbach and colleagues). The modified memory trace is then reconsolidated, stabilising these memories into long-term memory that can be remembered at later retrieval attempts. Whilst reconsolidation theory has been proposed to account for memory updating effects, direct evidence of reconsolidation modifying episodic memories in the human brain is lacking. Indeed, the majority of studies that monitor direct cellular changes to neural memory networks come from animal research (Hardt et al., 2010). What’s more, reconsolidation mechanisms are thought to be dependent on longer time delays between learning and retrieval events to accommodate the cellular changes within neural networks that update episodic memories. Furthermore, reconsolidation does not account for the updating of memories at much shorter time scales, suggesting that updating following reactivation can occur without reconsolidation mechanisms (Gisquet-Verrier & Riccio, 2012; Gisquet-Verrier et al., 2015).

An alternative theory to account for the updating effects seen by Hupbach and colleagues is the Temporal Context Model (Temporal Context Model; Howard & Kahana, 2002). This computational model suggests that items are bound to a mental, temporal context (defined as the running average of recently experienced items). By adapting the Temporal Context Model, Sederberg, Gershman, Poyn and Norman (2011) suggested that, in the Hupbach and colleagues studies, reminding participants of list 1 reactivates the temporal context associated with list 1 items. During list 2 presentation, items become bound to the reactivated list 1 context, meaning that participants then associate and misattribute items from list 2 to the context of list 1. No misattribution of list 1 items to list 2 contexts is observed because list 1 items are never bound to the list 2 context. Evidence of the concept that context reinstatement is critical to induce memory updating was shown in a study by Gershman, Schapiro, Hupbach and Norman (2013). In this study using fMRI pattern classification techniques, evidence of list 1 context reinstatement was observed ~2 seconds prior to the presentation of a list 2 object on screen only when objects from list 2 went on to be misattributed to list 1 in a subsequent recognition task. Such neural evidence for pre-stimulus list 1 context reinstatement also predicted increased confidence judgements during recognition when participants misattributed list 2 items to list 1.

Consequently, this study supports the Temporal Context Model theory postulating that the reactivation of contextual information allows the binding and updating of novel items to retrieved memories.

It is important to emphasise that neither the reconsolidation nor Temporal Context Model theories have been widely accepted to account for reactivation-induced memory updating effects. Furthermore, there is still debate regarding the precise mechanisms by which memory traces are stored in the brain that underlie reactivation-induced updating. Two ideas have been proposed; a storage-modification account and a retrieval-interference account. Storage modification is one of the assumptions of reconsolidation theory, whereby cellular changes occur to reactivated episodic memory traces that modify the neural representation of these memories (Hardt, Einarsson & Nadel, 2010). This modified neural trace is reconsolidated and stored in long-term memory that is accessed during subsequent retrieval attempts. However, it is difficult to directly observe the modification of reactivated memories in the human brain. As an alternative account, it has been suggested that memory updating effects results from a competition between separate mnemonic traces during retrieval. These retrieval competition accounts argue that the original memory trace is not modified during reactivation. Instead, a separate trace is formed during reactivation that represents the reactivated memory as well as information shown during the reactivation event (Riccio, Millin & Bogart, 2006). Consequently, at future retrieval attempts, there is selection competition between the memory for the original event and the memory of the reactivation event. The resolution of this retrieval competition depends on several factors, including the overlap of contextual information between learning, reactivation and subsequent retrieval events (which may be dependent on the overlap of temporal context), consistent with classic effects such as Tulving and Thomson’s (1973) encoding-specificity principle.

Supporting the retrieval competition theory, McCloskey and Zaragoza (1985) used an adapted version of the misinformation paradigm (Loftus, 2005), whereby participants initially viewed a slideshow depicting a crime. After this, participants read narratives of the crime, with details corresponding to the original event (control condition) or some details of original event modified (misled condition). A subsequent recognition task tested participants memory for the original versus misled information, with the classic finding that participants have poorer recognition accuracy for the original event in the misled condition (Loftus, 2005), taken as evidence of the original memory being updated by misinformation. However, in the McCloskey and Zaragoza (1985) study, a modified recognition test following the reading of narratives presented the original information alongside a completely new detail that was not shown during the narrative task. With this modified recognition task, participants in the control and misled conditions had similar recognition accuracy for the original event, indicating that participants could still access the original memory given optimal conditions during recognition tasks. Thus, the question regarding how episodic memories are modified by reactivation is far from resolved. These different theoretical accounts were considered during the design of studies, analysis of data and interpretation of results in the current research programme (see especially general discussion in Chapter 7).

1.2.1. Subjective retrieval processes and memory updating

An important factor that may influence episodic memory updating, addressed briefly in this introduction so far, is the role of subjective retrieval processes during retrieval attempts. Episodic memory retrieval can be dissociated by the objective accuracy of retrieval attempts (are retrieved memories correct or erroneous) and the subjective experience of retrieval (participants own judgement of retrieved information, regardless of the actual accuracy). In the context of recognition memory (relevant for the present thesis), subjective retrieval processes can be measured using several methods, including procedures such as ‘Remember/Know’ (Tulving, 1985) where participants either ‘remember’ specific encoding-related details of a memory (reflecting recollection related processes), or just ‘know’ that a memory has been encountered before (reflecting familiarity related processes), regardless of whether the retrieved information is accurate to the original event.

Furthermore, the contents of retrieval, based on either familiarity or recollection, can be judged according to confidence judgements (i.e. how confident an individual is that what they remembered is correct). Theoretically, confidence judgements associated with familiarity can be explained according to signal detection models, such that the strength of familiarity judgements determines the subjective appraisal of memories, in accordance with an individual’s response criterion (Busey, Tunnicliff, Loftus & Loftus, 2000; Parks & Yonelinas, 2009; Wixted & Mickes 2010). However, regarding recollection, it has been argued that recollection judgements are associated with high confidence judgements only, on the basis that recollection is an ‘all-or-none’ process where we either recollect details of a memory or not (Yonelinas et al., 1996). However in recent years, recollection has been shown to reflect a ‘some-or-none’ process (Harlow & Donaldson, 2013; Harlow & Yonelinas, 2016), such that the ‘precision’ of recollected memories can vary, should recollection occur, meaning that subjective appraisal of recollected information could also vary according to how accurate recollection is. Consequently, the relationship between objective and subjective processes reflects a complex interplay between the true contents of memory and how we appraise retrieved information.

Alternatively, subjective retrieval can measured in terms of the confidence/vividness of a retrieved memory. A simple interpretation suggests that objective and subjective retrieval processes are two sides of the same coin, such that accurate retrieval should be associated with elevated confidence/vividness whereas inaccurate retrieval should be associated with reduced confidence/vividness. Indeed, neuroimaging studies show increased confidence during successful memory retrieval is associated with increased neural measures of retrieval success, demonstrated with fMRI (Thakral, Wang & Rugg, 2015; Yonelinas, Otten, Shaw & Rugg, 2005) and EEG (Woodruff, Hayama & Rugg, 2006).

However, other lines of research indicate that the relationship between objective and subjective retrieval processes is more complex. Indeed, the neural correlates of objective and subjective retrieval processes have been dissociated with evidence from fMRI (Richter, Cooper, Bays & Simons, 2016) and EEG (Rutishauser, Aflalo, Rosario, Pouratian & Andersen, 2018; Woodruff, Hayama & Rugg, 2005). What's more, Kim and Cabeza (2007) showed that elevated confidence during false recognition attempts was associated with increased activity in prefrontal and parietal cortices, regions otherwise implicated in successful episodic memory retrieval (Rugg & Vilberg, 2013). Consequently, it is over simplistic to assume that objective and subjective retrieval mechanisms are completely overlapping.

As a more direct study of the interaction between objective and subjective retrieval processes, a collection of studies by Roediger and DeSoto (DeSoto & Roediger, 2014; Roediger & DeSoto, 2014) succinctly reflect how the design of retrieval tasks can influence the relationship between objective and subjective retrieval processes. Using an old/new recognition task, participants first encoded word lists during study, with the same words shown during a subsequent recognition test. During test, participants also saw new words that were either semantically related to old words (related lures) or not semantically related to old words (unrelated lures). Confidence responses during test were acquired for recognition responses to all item types. Both DeSoto and Roediger (2014) and Roediger and DeSoto (2014) established that, for old and unrelated lures, positive correlations between accuracy and confidence were found such that correct recognition correlated with higher confidence judgements. However, for related lures, negative or no correlations between accuracy and confidence was shown, driven by elevated confidence responses for the false recognition of related lures. Consequently, these results suggest that the relationship between confidence and accuracy is dependent on the similarity of old and new items, an important consideration for the present thesis where recognition tasks presented face stimuli that shared some perceptual features (see Chapter 2, section 2.1).

Confidence judgements during recognition can be viewed as representing the strength of recognition signals evoked during retrieval. Signal detection theories of recognition memory (Busey et al., 2000; Parks & Yonelinas, 2009; Thakral, Wang & Rugg, 2015; Wixted & Mickes 2010) suggest that stronger recognition signals result in higher confidence responses. This goes not only for correct recognition attempts, but also for incorrect recognition of items that were similar to true memories (such as the related lures in DeSoto & Roediger, 2014; Roediger & DeSoto, 2014). Furthermore, since elevated subjective retrieval judgements correlate with enhanced neural reactivation (Johnson, McDuff, Rugg and Norman, 2009; Kuhl & Chun, 2014; Thakral, Wang & Rugg, 2015), this suggests that confidence during retrieval may also predict the extent to which memories become updated, mediated by reactivation mechanisms that may be crucial for episodic memory updating to occur (see section 1.2).

Direct evidence that subjective retrieval processes predict the updating of episodic memories has been found in several studies by St. Jacques and colleagues (St. Jacques, Olm & Schacter, 2013; St. Jacques & Schacter, 2013). In these studies, participants were asked to take part in a museum tour, following which participants completed an initial recognition task. Here, photographs of the original tour were presented, with participants providing ratings on ‘reliving’ (higher reliving corresponded to increase subjective retrieval experience). Following reliving judgements, participants were shown a lure photograph depicting information that was not present in the original museum tour. In a subsequent recognition task, participants provided old/new judgements for the original and lure photographs, as well as new photographs not shown previously. In both studies, it was found that greater reliving for old items during the first recognition task increased the likelihood that these photographs would be correctly recognised during subsequent recognition. However, greater memory reliving during initial recognition also increased the likelihood that lure photographs would be falsely recognised during subsequent memory. Furthermore, St. Jacques, Olm and Schater (2013) found that, for memories accompanied with higher reliving, BOLD signals from fMRI were larger in regions associated with episodic memory reactivation and contextual reinstatement, such as parahippocampal gyrus, retrospenial cortex and inferior parietal cortex. In addition, during initial recognition, activity in regions implicated in the incorporation of novel information into episodic memories, such as the ventromedial prefrontal and anterior hippocampus, predicted that participants would falsely recognise lure photographs during subsequent recognition. These finding thus show novel evidence of the brain networks that are involved in memory updating, and how they relate to subjective retrieval experience and objective memory accuracy.

The reviewed literature suggests that subjective retrieval processes may be related to enhanced updating of episodic memories. Specifically, subjective retrieval processes may be associated with increased reactivation of memories during retrieval that can then become updated by information present during the retrieval event, consistent with the framework posited by Hupbach and colleagues (Hupbach et al., 2007; 2008). A prediction that follows from these experiments is that active retrieval attempts and subjective experience of recognition is critical for the updating of face memories, an assumption that was tested in the experiments of Chapter 3 using recognition confidence as a measure of subjective retrieval.

Confidence was used as a measure of subjective retrieval in the current thesis as it was decided that participants would be better at evaluating their recognition confidence better than subjective retrieval processes used in previous research (e.g. vividness or reliving judgements used in St. Jacques et al., 2013), given the minimal information participants were able to remember about the target face memories. A consequence of this decision, however, questions whether confidence is equivalent to other subjective retrieval factors. No such claim can be made, given previous research showing divergences in the neural mechanisms associated with different objective/subjective retrieval processes (Richter et al., 2016). However, given that previous research has suggested an association between enhanced reactivation with increased confidence judgements (Thakral et al., 2015; Woodruff et al., 2006), it was justified to use confidence as a possible mechanism that could modulate face memory updating.

1.2.2. Is retrieval necessary to induce updating?

A second factor examined in this thesis assessed whether active retrieval attempts modulate face memory updating. The reviewed literature so far assumed that reactivation of memories is critical to induce updating. However, as discussed in the opening paragraphs of this chapter, retrieval attempts of previously encoded information are thought to be a powerful mechanism to induce memory updating compared to tasks where information is re-studied (Butler et al., 2006; Marsh et al., 2007; Roediger & Butler, 2011; Roediger & Marsh, 2005). An important study by Bridge and Voss (2014) addressed the role of active retrieval attempts in memory updating. Their participants completed four blocks of learning, refresh and final recognition tasks. During learning, participants encoded several object-location associations on a 2-D screen. During the subsequent refresh task, participants completed an ‘active’ retrieval (in block one & three) or a ‘passive’ re-study task (in block two & four). The active retrieval task presented participants with an object in the centre of the screen, and participants were asked to retrieve the location associated with the object from learning and to place objects at the retrieved location. In contrast, the passive re-study task required participants to place objects at a location that was indicated by a yellow box. The location of the yellow box was determined by the locations that participants chose during one of the previous active refresh trials. Therefore, the divergence of object-locations from the original location to the refresh location was matched between active and passive refresh, ensuring that retrieval accuracy was matched between the two conditions. This task design thus addresses a common issue of retrieval practice studies, where the retrieval and re-study tasks are not necessarily matched in terms of the degree to which they reinforce accurate information.

Following both refresh tasks, participants completed a recognition test that presented the objects placed at one of three different locations; the location originally shown at learning, the location selected during refresh and a control location equidistant between the other two locations. Recognition performance showed that, following the active retrieval task, participants were more likely to select the refresh location compared to original or control locations. In contrast, following the passive re-study task, participants were more likely to select the original location compared to refresh or control locations. Hippocampal activity predicted updating during both refresh conditions, however activity during active refresh predicted whether the retrieved object location would be associated with the original context (predicting memory updating), whereas activity during passive refresh predicted whether the original object location would be associated with context at refresh, preserving the original memory location to be correctly recognised later.

Whilst the study from Bridge and Voss (2014) did not necessarily measure the extent to which memories are reactivated during active and passive refresh tasks (rather, this was just assumed to occur in these tasks), the results suggest that active retrieval attempts may also enhance the updating of face memories compared to re-study. This issue was addressed in the experiments in Chapter 4 of the present thesis, with Experiments 4a-c using a similar design to that used by Bridge and Voss (2014). This involved comparing active retrieval and restudy of face memories whilst matching error between retrieval/re-study conditions, but with using images derived from face space models (see Section 1.5).

1.3. Electrophysiological correlates of episodic memory

In summary, this literature review has thus far discussed the role of reactivation, subjective experience and active retrieval attempts in the updating of episodic memories. Whereas different theoretical accounts of how retrieval-induced updating occurs are still debated, there is evidence from behavioural data indicating that memory reactivation is critical to induce the updating of episodic memories, and that encoding mechanisms engaged during memory reactivation are engaged to update episodic memories. However, the behavioural results discussed so far are not conclusive with regards to the underlying mechanisms that give rise to memory updating. Instead, neuroimaging data may be necessary to provide more decisive evidence. Whilst fMRI methods have been used in prior literature to further our understanding of the neurocognitive dynamics of memory updating, this thesis will focus on the use of electrophysiological techniques, specifically EEG, as a tool to understand how different processes interact to induce memory updating. EEG measures the electrical activity of the brain with a high temporal resolution

(Kappenman & Luck, 2012), allowing a critical examination of the order of cognitive operations that occur during the updating of episodic memories, and may enable a dissociation of different component operations. Given that EEG is the main neuroimaging method used for the present thesis, literature related to EEG studies of episodic memory, including the use of ERP and time-frequency analysis, will be discussed in the next sections. These two techniques offer distinct insights into the neural mechanisms of episodic memory for several reasons. Firstly, oscillations can either be evoked or induced to stimulus/response onsets, i.e., evoked oscillations are ‘phase-locked’ to the onset of stimuli, while induced activity is not phase-locked (David, Kilner & Friston, 2006). For this reason, induced oscillatory effects will vary across trials in their temporal onset (as the onset of the cognitive process associated with these oscillations will also vary). Consequently, averaging trials during the computation of ERPs will average out any induced oscillatory effects (Bastiaansen et al., 2011), meaning that there may important differences between what ERP and oscillatory analyses can show with regards to episodic memory processing. A further critical discussion of the strengths and weaknesses of these two approaches will be provided in Chapter 2 (section 2.4). Instead, the current literature review will focus on summarising research using ERP and time-frequency techniques in the context of episodic memory encoding, retrieval and updating.

1.3.1. ERP correlates of episodic memory

The ERP technique has been used to study the neural basis of cognition since the 1970’s (Kappenman & Luck, 2012), and has been a widely used tool to understand the neurocognitive mechanisms associated with episodic memory encoding, retrieval and updating (Wilding & Ranganath, 2011). The ERP correlates of episodic memory retrieval have received much attention over recent decades. Particularly, the ERP method has been used to resolve theoretical debates of how recognition memory systems operate during retrieval, especially the issue of whether recognition is served by a single process (Wixted, 2007) or dual processes of familiarity and recollection (Yonelinas, 2002). The balance of evidence from this ERP literature suggests that recognition memory is supported by two processes (Rugg & Curran, 2007), with two distinct ERP measures separately associated with familiarity and recollection, the FN400 and the left parietal old/new effect (LPE). The FN400 has been often associated with familiarity-related recognition (a quantitative estimate of having seen an item previously). This ERP, located at mid-frontal sites that peaks at ~300-500ms from stimulus onset, is generally more positive for hits (correctly recognising old items) compared to misses (failing to recognise old items), correct rejections (correctly identifying new items as new) and false alarms (incorrect recognising new items as old; Wilding & Ranganath, 2011). Interestingly, increased FN400 positivities are seen for false alarms of items that were semantically related to items shown during learning (Curran, 2000). Furthermore, the FN400 effect has been shown to correlate with recognition confidence, such that FN400 effects are largest for high confidence hits, followed by low confidence hits, low confidence correct rejections and smallest for high confidence correct rejections (Addante, Ranganath & Yonelinas, 2012; Woodruff, Hayama & Rugg, 2006; Woroch & Gonsalves, 2010). These results have led to the conclusion that the FN400 reflects the signal match evoked between items shown during recognition and of internal mnemonic representations, linking this ERP correlate with familiarity-based recognition judgements. However, it must be noted that alternative accounts suggest the FN400 may reflect priming, an implicit memory process that facilitates memory judgements without conscious awareness of having encountered information before (see Bridger, Bader, Kriukova, Unger & Mecklinger, 2012; Paller, Voss & Boehm, 2007; Rugg & Curran, 2007; Voss & Paller, 2008; Wilding & Ranganath, 2011 for discussions).

In contrast, recollection-related retrieval processes (remembering qualitative details of an encoding event) have been associated with the LPE. The LPE has a prominent left parietal topography that is more positive when participants can recollect contextual information of a previous episode, peaking between ~500-800ms after stimulus onset (Wilding & Ranganath, 2011). Furthermore, studies have shown the left parietal old/new ERP to be increased when retrieval attempts contain correct, highly confident recollections, compared to low confidence but correct recollections and correct rejections, emphasising a dissociation between the FN400 and left parietal old/new effects with recognition confidence (Addante et al., 2012; Woodruff et al., 2006; Woroch & Gonalves, 2010). The left parietal old/new effect has been suggested to reflect activity in the parietal cortex, a region that has been consistently linked to episodic memory retrieval. In recent years, several theoretical arguments have been proposed to account for the function role of parietal cortex in episodic memory, including attention-to-memory (Cabeza, Ciaramelli, Olson & Moscovitch, 2008) and episodic memory buffer hypotheses (Vilberg & Rugg, 2008). However, recent evidence suggests that the parietal cortex supports episodic memory processing by binding information from multiple areas of cortex (Shimamura, 2011), allowing individuals to make explicit retrieval decisions (recollection, enhanced subjectivity) on the content of retrieved episodic memories (Ally, Simons, McKeever, Peers & Budson, 2008; Vilberg & Rugg, 2009). However, it should be emphasised that EEG methods do not allow firm conclusions to be drawn regarding generators of scalp ERP effects (Luck, 2005), and therefore these links with the parietal cortex are only tentative.

The FN400 and LPE ERPs are often found during recognition tests for stimuli such as words, objects or scenes that contain rich semantic/conceptual information. However, it has been questioned whether these ERP components reflects general memory retrieval processes, or instead are specific to the type of material that is being retrieved. For example, several studies have established that the ERP correlates for the recognition of faces may differ to the FN400 and LPE. Face recognition has been associated with a number of ERP correlates that index different processing stages/cognitive operations, including the N170 (Bentin et al., 1996; Eimer, 2011) and N250 (Schweinberger, Huddy, & Burton, 2004; Schweinberger, Pfütze, & Sommer, 1995). It has been suggested that the N170 reflects the perceptual identification of faces (compared to non-face stimuli; Itier & Taylor, 2004). On the other hand, the N250 has been linked with activation of unique facial representations, particularly for familiar faces that have an established mnemonic representation (Schweinberger, Huddy, & Burton, 2004). However, the N250 has also been shown for faces that were learnt during one experimental session (Pierce et al., 2011; Tanaka, Curran, Porterfield, & Collins, 2006), suggesting that this ERP effect can be found immediately after the development of stable face representations (Andrews, Burton, Schweinberger & Wiese, 2017). Electrophysiological correlates of face processing have been observed even earlier than the N170 (e.g. Herrmann, Ehlis, Ellgring, & Fallgatter, 2005; Nemrodov, Niemeier, Mok & Nestor, 2016), and ERPs also distinguish highly familiar and unknown faces well beyond 250ms (Wiese et al., 2019). Thus, there is still some debate regarding the functional properties of these various ERP effects in relation to the various stages of face processing.

Whilst the aforementioned ERP studies of face processing have identified several correlates of the stages of face recognition, these ERPs reflect the detection and identification of a stable facial representation rather than mechanisms of retrieving an episodic memory of a face. To examine the ERP correlates of familiarity and recollection of face memories, Mackenzie, Donaldson and colleagues (2007, 2009, 2018) tested participant’s memory for novel faces with the old/new recognition paradigm. These faces were novel as participants had never seen them prior to the experiment, thus should have minimal semantic or conceptual information stored in long-term memory. Interestingly, Mackenzie and Donaldson (2007) showed that the ERP correlate of familiarity decisions for these faces was associated with a posterior positivity from ~300-500ms, in contrast to the FN400 effect that has a more anterior topography. Furthermore, recollection judgements were associated with positive ERPs from ~500-900ms at anterior electrodes, in contrast to the recollection of famous faces (Mackenzie, Alexandrou, Hancock & Donaldson, 2018) and names associated with faces (Mackenzie & Donaldson, 2009) that showed the typical LPE. In sum, these studies suggest that the ERP correlates of recognition memory are dependent on the semantic status of episodic memories (Nie et al., 2014), an important consideration for the experiments of this thesis that used novel face stimuli to study the updating of item memories (see Chapter 2, section 2.1).

ERPs have also been examined during episodic memory encoding to compare the neural activity during the presentation of items that will be subsequently remembered compared to items that will be subsequently forgotten in a future retrieval task. This ‘subsequent memory effect’ (Paller & Wagner, 2002) distinguishes the neurocognitive mechanisms contributing to successful versus unsuccessful memory encoding. A frequent finding is that ERPs for memories that will be successfully remembered are more positive than ERPs for memories that will be subsequently forgotten (Otten & Rugg 2001), though negative ERP deflections have been observed during encoding of memories with little semantic information (Otten, Sveen & Quayle, 2007). These results suggest that the nature of to-be-remembered information influences the neural mechanisms associated with the encoding of different items.

Research has also examined ERP components to understand the neurocognitive mechanisms of successful memory encoding. For example, a study by Griffin, DeWolf, Keinath, Liu and Reder (2013) analysed the FN400 and LPE ERPs during a study phase where objects were presented once or twice. Objects presented twice were either identical repetitions (the exact same picture of an object was repeated) or conceptual repetitions (a different exemplar of the object shown each time). ERP analyses during study showed that the FN400, despite being larger during the second presentation of objects, did not predict subsequent memory accuracy. In contrast, the LPE was larger for objects in the conceptual repetition condition, which predicted whether participants would successfully recollect the condition of the objects. The authors suggest that participants created a durable and strong episodic memory during encoding by retrieving previous memory associations, akin to the retrieval practice effect (Roediger & Butler, 2011). Performance during retrieval may have been facilitated by pattern separation mechanisms during encoding to ensure that unique mnemonic representations were formed for the overlapping objects, that could then be successfully recollected at retrieval (Yassa & Stark, 2011).

More recently, a study by Kamp, Bader & Mecklinger (2017) suggests that the timing and topography of subsequent memory ERP effects may depend on the nature of encoding tasks, and participant strategies used during encoding. Specifically, the authors suggest that two subsequent memory ERPs, a ‘P300’ ERP at parietal sites and a slow-wave ERP at frontal electrodes, may index encoding of items and item-context relations, respectively. In their study, Kamp et al. (2017) asked participants to encode word pairs during a study phase in a ‘definition’ condition (word pairs were ‘unitised’ as a single item) or a ‘sentence’ condition (word pairs form inter-item contextual associations). During the study phase, ERP analyses showed the P300 component predicted subsequent retrieval accuracy for word pairs in the definition but not sentence condition, consistent with the view that this component reflects the detection and recognition of salient items (Polich & Kok, 1995), suggesting participants may have attended to these items more to boost encoding (Ciaramelli, O'Grady & Moscovitch, 2008). In contrast, frontal slow-wave ERPs predicted subsequent retrieval accuracy in both conditions, with the authors interpreting that working-memory mechanisms, potentially attributed to pre-frontal cortex (Blumenfeld & Ranganath, 2006), were present during both conditions to facilitate successful encoding. Consequently, ERPs during initial stimulus presentation that predict successful memory encoding, indexed by participants ability to subsequently remember these stimuli, reflect a variety of cognitive mechanisms that may be task and stimulus dependent.

The ERP correlates of episodic memory encoding and retrieval have only been used a few times to investigate episodic memory updating. In one study by Bridge and Paller (2012), participants first learnt several 2-D object-location associations on-screen before completing multiple retrieval tests. During test 1, participants were cued with objects in the centre of the screen, and were asked to retrieve the location associated with objects during the previous learning phase, and place objects at this location. The same retrieval task was completed in a subsequent test 2. This paradigm allowed the authors to measure retrieval accuracy on a continuous scale, that being the distance (cm) between the original object location during study, and the location of objects placed during both test 1 and test 2. Focusing on the neural activity during test 1, ERPs were contrasted by retrieval success (was the object placed closer or further from the study location). Also, ERPs were contrasted by ‘future’ retrieval bias, this being whether participants retrieved a similar location in the subsequent test 2 (‘close’ retrieval bias) or retrieved a dissimilar location in the subsequent test 2 (‘far’ retrieval bias), akin to the subsequent memory effect (Paller & Wagner, 2002). This latter contrast was hypothesised to reveal the ERP correlates of memory updating mechanisms that were responsible for biasing future memories by incorporating errors made during retrieval attempts.

For retrieval success, ERPs were found to be more positive at frontal sites from 400-700ms when objects were placed closer to the study location. In contrast, ERPs for the close bias condition were more positive going from 700-1000ms at frontal and parietal electrodes compared to ERPs for the far bias condition. These findings indicate that reactivation of episodic memories (from 400-700ms) was followed by post-retrieval encoding mechanisms (700-1000ms) that update retrieved memories with content remembered during retrieval. However, the design of the Bridge and Paller (2012) study meant that, during both tests, participants responses always deviated from the original location (i.e., all responses were erroneous). Consequently, this study could not establish the neural correlates of the positive consequences of updating that establishes stable memory representations by strengthening accurate memories (a.k.a the retrieve-practice effect; see section 1.2.2). In addition, ERPs during test 2 were not analysed, thus the neural correlates of the consequences of retrieval-induced updating were not considered.

Building on the Bridge and Paller (2012) study, an ERP experiment by Liu, Tan and Reder (2018) examined both the positive and negative consequences of repeated retrieval. In their study, participants were asked to encode word pairs in an initial learning phase. Participants then completed a cued-recall test for word pairs in test 1, with the same cued-recall task repeated twice more throughout the study. For the ERP analysis, trials were categorised during test 1 according to whether participants' memory for word pairs was correct or incorrect. These conditions were then further categorised based on whether the word pair was correctly or incorrectly retrieved during the subsequent retrieval tests, creating three conditions; correct-correct, correct-incorrect and incorrect-incorrect.

Analysis of neural activity during test 1 showed that, in the 400-700ms time window, ERPs were more positive for the correct-correct condition compared to the two other conditions. Further, the correct-incorrect condition had significantly more positive

ERP’s than the incorrect-incorrect condition. However, during the 700-1000ms time window, ERP’s were more positive for the correct-correct condition compared to the other two conditions, with no difference found between correct-incorrect and incorrect-incorrect. Similar to Bridge and Paller (2012), these results suggest that, during the 400-700ms time window reactivation processes were engaged for successfully recalled word pairs. The later 700-1000ms time window was instead associated with encoding mechanisms that strengthened reactivated memories and therefore predicted subsequent retrieval accuracy during the repeated cued-recall test.

Further ERP analysis from Liu, Tan and Reder (2018) also analysed activity during the second test for trials where participants had already correctly retrieved word pairs during the previous test 1. During the second test, correct-correct, correct-incorrect and incorrect-incorrect conditions were again created according to retrieval accuracy during the current test 2, and subsequent test 3. Similar ERP effects were found to those in test 1 when analysing ERPs from 400-700ms, however, no ERP differences were found from 700-1000ms. The lack of ERP effects during the later time window was attributed to that fact that encoding mechanisms were not engaged during repeated correct retrieval, as these memories had already been successfully recalled during test 1, therefore there was no need for these established episodes to become updated during later testing.

In summary, the current section has reviewed the ERP literature with regards to the encoding, retrieval and updating of episodic memories. The ERP technique has played a critical role over recent decades to understand the neurocognitive mechanisms of these episodic memory processes. This literature will be returned to in Chapter 5, where the ERP correlates of the retrieval and updating of face memories was assessed.

1.3.2. Oscillatory correlates of episodic memory

Research into the oscillatory correlates of cognitive function, including episodic memory, have become more prominent since the 1990’s. EEG oscillations are rhythmic patterns of activity generated by populations of neural ensembles that can be measured from scalp electrodes (Bastiaansen, Mazaheri & Jensen, 2011). Neurons within an ensemble can either have firing rates that occur at the same time as other neurons within an ensemble, i.e. they synchronise, which leads to oscillatory power increases, or have neuronal firing rates that occur at different times to other neurons in the ensemble, i.e. they desynchronise, which leads to oscillatory power decreases. Oscillations are characterised by their frequency of cycles per second, with delta (1-4Hz), theta (4-8Hz), alpha (8-12Hz), beta (12-30Hz) and gamma (30-100Hz) oscillations thought to be related to functionally dissociable processes. As will be discussed in the following review, synchronisation of theta and gamma oscillations are typically related to successful episodic memory encoding and retrieval. In contrast, desynchronisation of the alpha and beta frequency bands are most often associated with successful memory encoding and retrieval. This synchronisation/desynchronisation ‘conundrum’ (Hanslmayr, Staresina & Bowman, 2015) suggests a dissociation in the functional properties of these frequency bands with respect to episodic memory encoding, retrieval and updating.

Focusing on theta oscillations, synchronisation within this frequency band has been consistently linked to various stages of episodic memory processing. During episodic memory encoding, theta power increases are found during the presentation of items that will subsequently be remembered compared to items that will be forgotten during later retrieval tests (Hanslmayr et al., 2011; Osipova et al., 2006). Hanslmayr et al. (2011) localised the source of this theta-power subsequent memory effect to the parahippocampal regions, suggesting that theta power generated from regions in the medial temporal lobe promotes the encoding of contextual information during episodic memory formation to facilitate subsequent remembering (Hanslmayr & Staudigl, 2014). In addition, fronto-central theta power effects have been found during encoding of items that are subsequently remembered with accurate recollection judgements (Friese et al., 2013), and with elevated retrieval confidence (Wynn, Daselaar, Kessels, & Schutter, 2019). The frontal topography possibly reflects top-down goal control processes that prioritise the encoding of goal-relevant items into long-term memory (Blumenfeld & Ranganath, 2007). Similar findings have been established in the working memory literature, whereby frontal theta power increases correlate with maintaining or rehearsing items in working memory (Khader & Rösler, 2011; Meeuwissen, Takashima, Fernández & Jensen, 2010), supporting the idea that the frontal distribution of the theta effect relates to control processes during formation of episodic memories.

Theta power increases have also been associated with successful episodic memory retrieval. With scalp recorded EEG, such effects have been observed primarily at frontal (Klimesch, Doppelmayr, Schimke & Ripper, 1997) and left parietal locations (Jacobs, Hwang, Curran & Kahana, 2006; Klimesch et al., 2000). Similar frontal theta effects has also been demonstrated for successfully retrieving target memories amongst competing, interfering stimuli (Waldhauser, Johansson & Hanslmayr, 2012). These results suggest that frontal theta power mediates controlled, top-down mechanisms when retrieval requires controlled, goal-directed selection during memory search. On a separate note, studies have established increased left parietal theta power during successful retrieval in cued recall (Hanslmayr et al., 2011) and source memory tasks (Addante et al., 2011), and for high confidence retrieval judgements (Wynn et al., 2019). Given the functional and topographical similarity of the parietal theta effect to the left parietal ERP effect, this oscillatory correlate may correspond to recollection-related retrieval processes.

To summarise, theta oscillations have been associated with successful episodic memory encoding and retrieval across a large number of studies. Furthermore, activity at theta frequencies is thought to functionally couple with higher frequency gamma oscillations, such that gamma power bursts occurring at specific phases of the theta cycle supports episodic memory encoding (Friese et al., 2013; Hanslmayr & Staudigl, 2014; Lisman 2010) and retrieval (Axmacher, et al., 2010; Fuentemilla, 2018; Hanslmayr & Staudigl, 2014; Kerrén, Linde-Domingo, Hanslmayr & Wimber, 2018; Köster et al., 2014; Lega, Burke, Jacobs & Kahana, 2014).

In contrast to theta and gamma frequencies that typically increase in power during successful encoding and retrieval, alpha and beta desynchronisations are often found to correlate with successful episodic memory processes. For example, alpha/beta power reductions are observed during presentation of items during encoding that are successfully remembered in a subsequent retrieval task (Friese et al., 2013; Hanslmayr, Staudigl & Fellner, 2012; Klimesch et al., 1997). Furthermore, Hanslmayr, Spitzer and Bauml (2008) observed alpha/beta desynchronisation to be dependent on the ‘levels of processing’ of memories (Craik & Lockhart, 1972), when encoding tasks required participants to make ‘deep’ judgements on items, such as the semantic content of items, as opposed to ‘shallow’ judgements that are based on perceptual features. Based on converging evidence, beta desynchronisation has been linked to inferior prefrontal cortex (Hanslmayr et al., 2011), a region implicated in successful semantic memory encoding (Otten & Rugg, 2001). Thus, the results in these studies are consistent with alpha/beta desynchronisation effects reflecting access to semantic information during encoding, which in turn facilitates the formation of episodic memories.

A large body of evidence suggests that increased alpha/beta oscillatory power reflects an inhibitory mechanism in the brain (Klimesch, Sauseng & Hanslmayr, 2007;

Jensen & Mazaheri, 2010). Therefore, desynchronisation of alpha/beta oscillations may reflect a ‘release from inhibition’, allowing information to be successfully accessed and transferred between brain regions during episodic memory encoding. For example, in studies from Hanslmayr et al. (2009) and Hanslmayr et al. (2011), alpha/beta desynchronisation during memory encoding may have allowed the transfer and incorporation of semantic information into episodic memory (Hanslmayr, Staresina & Bowman, 2016; Parish, Hanslmayr & Bowman, 2016). These studies collectively suggest that reduced inhibition of task-relevant brain regions during encoding, mediated by alpha/beta oscillations, facilities episodic memory formation.

Whereas the majority of studies link alpha/beta desynchronisation with successful memory encoding, some studies have suggested that alpha/beta synchronisation can also predict successful memory formation. For example, posterior alpha power increases have been correlated with successfully maintaining and rehearsing items in working memory for items that would be subsequently recognised vs. forgotten (Khader & Rösler, 2011; Meeuwissen et al., 2010; Poch, Valdivia, Capilla, Hinojosa & Campo, 2018). Furthermore, Bonnefond and Jensen (2012) conducted a study where participants were initially presented with a series of letter digits, after which a relevant (different letter) or irrelevant (symbol) was presented. Participants were then presented with a probe and asked to identify whether the probe was shown in the original list. During distractor presentation, alpha power increases at posterior electrodes was shown during presentation of relevant versus irrelevant distractors, with alpha power increases correlating with participants ability to suppress distracting information. Overall, these studies indicate that during maintenance/rehearsal of items in working memory, alpha/beta synchronisations serve to inhibit brain activity that would otherwise interrupt successful memory formation.

The conflicting nature of alpha/beta (de)synchronsiation during memory encoding and rehearsal has been further clarified in studies that directly compared these encoding and working memory processes during episodic memory formation. For example, Babu Henry Samuel et al. (2018) showed that alpha power decreases correlated with increased digit set size during the presentation of digits on-screen. On the other hand, enhanced alpha power correlated with increased digit set size during rehearsal of items following digit off-set. Thus, during presentation of information to-be-encoded, inhibitory mechanisms are disengaged in task-relevant brain regions to promote successful encoding of items in short-term memory. Following this, during periods of retention, inhibitory processes are engaged in regions of the brain that process competing memories, promoting the successful storage of items into long-term memory for subsequent retrieval.

Alpha/beta desynchronisation effects have also been shown to correlate with successful episodic memory retrieval. Similar to encoding, alpha/beta desynchronisation during retrieval is thought to reflect the release of inhibition in material specific brain regions that facilitate access to stored episodic memory representations (Graetz, Daume, Friese & Gruber, 2018; Hanslmayr, Staresina & Bowman, 2016; Parish, Hanslmayr & Bowman, 2016). For example, Khader et al. (2010) asked participants to encode words that were paired with either a 2-D location or an object during a learning phase. During retrieval, participants were then cued with words and recalled the associated location or object. During successful recall of locations, alpha desynchronisation was observed at posterior electrodes, whereas successful recall of objects was associated with alpha desynchronisation at frontal electrodes. The spatial divergence of alpha desynchronisation effects supports the argument that this oscillatory mechanism facilities retrieval by disinhibiting task-relevant brain regions during retrieval, since task relevant regions might be different for different types of memory features (e.g. locations vs. objects).

Alpha/beta (de)synchronisation has also been found to be associated with retrieval of target information in the presence of competing but irrelevant memories. For example, Waldhauser, Johansson and Hanslmayr (2012) conducted a study where participants initially encoded object-colour associations during study. In this task, objects were shown in two cycles, with each cycle associating the object with a coloured square that was shown either in the left or right visual hemifield. Squares were presented in the same colour in both hemifields (non-interference condition) or presented with different colours in each hemifield (interference condition). During retrieval, objects were shown with a blank square in either the left or right hemifield, with participants being asked to retrieve the target colour association (shown in cycle one) instead of the competitor colour association (shown in cycle two). In the interference condition, early alpha/beta power increases were observed at posterior electrodes on the hemisphere contralateral to the competitor colour association. Following this early effect, alpha/beta power decreases were seen at posterior electrodes on the opposite hemisphere (contralateral to the target colour association). This study shows further evidence that alpha/beta oscillations appear to facilitate successful remembering by inhibiting activity in irrelevant brain regions and releasing inhibition of neural activity in relevant brain regions that contain targeted episodic memories.

Further studies have also examined how alpha/beta oscillations correlate with the updating of episodic memories. First, Guran, Herweg & Bunzeck (2019) applied oscillatory analysis to the retrieval-practice paradigm (see section 1.2.2). In this study, young and old participants initially encoded several scenes before completing a retrieval task (were scenes old or new) or a re-study task (were scenes indoor or outdoor). A subsequent recognition task was completed 10 minutes or 24 hours later, which presented scenes from either the study and retrieval/re-study phase (old scenes) or were not shown previously (new scenes). Behaviourally, recognition performance was better for scenes in the retrieval condition versus re-study condition (consistent with the retrieval-practice effect). It was also shown that alpha/beta desynchronisations were observed during the retrieval versus re-study task, with larger alpha power desynchronisation correlated with improved later recognition accuracy. These results suggest that alpha/beta desynchronisation not only correlates with successful episodic memory retrieval, but is shown to be specific to retrieval (compared to re-study tasks) that predict retrieval-induced enhancements of episodic memories.

An additional study examining the oscillatory correlates of memory updating, by Bäuml, Hanslmayr, Pastötter & Klimesch (2008), asked participants to initially encode a list of words during study. Participants were then told to either ‘remember’ or ‘forget’ list 1 prior to encoding a second word list. A subsequent retrieval task asked participants to freely recall words from both lists. In the remember condition, participants had significantly better free recall performance for list 1 versus list 2 words, whereas participants in the forget condition had better recall performance for words in list 2 versus list 1. From the oscillatory data, participants in the forget condition showed increased alpha power at temporal electrodes during the presentation of list 2 that predicted better memory for list 2 items during recall. This increased alpha power effect was thought to reflect an active inhibition of memories from list 1 in order to optimise encoding of memories in list 2, possibly reflecting a mechanism of episodic memory updating.

In summary, the oscillatory correlates of episodic memory encoding, retrieval and updating generally separate oscillatory mechanisms that allow for the formation and retrieval of coherent episodes (represented by theta and gamma), supported by alpha/beta oscillatory processes to inhibit/disinhibit irrelevant or relevant brain regions to facilitate encoding/retrieval processes. These mechanisms will be considered again in Chapter 6, where oscillatory mechanisms underlying the objective, subjective and updating mechanisms of face memories were examined.

1.4 Recognition-induced updating of item memories

As summarised, previous research has demonstrated the behavioural and neural mechanisms associated with the retrieval-induced updating of episodic memories. The key arguments to be emphasised are; 1) reactivation of memories is critical for memories to become updated (Bridge & Paller, 2012), 2) enhanced subjective experience of retrieval increases the likelihood of memory updating (St. Jacques et al., 2013), and 3) active retrieval attempts enhance updating compared to re-study (Bridge & Voss, 2014). These studies tested the updating of elaborate episodic memories, including associative (object-location associations) and autobiographical memories. Furthermore, the studies by Bridge and Paller (2012) and Bridge and Voss (2014) tested the updating of memories via cued-recall retrieval tasks. Whilst these studies developed our understanding of the mechanisms of retrieval-induced updating, it is unclear as to how other types of memories, such as item memories, may become updated via recognition-based retrieval mechanisms.

It is important that the mechanisms of recognition-induced updating of item memories are studied as neurocognitive differences have been established between item recognition and elaborate memory recall. According to Ranganath and Ritchey (2012), item-based recognition memory judgments, particularly those made based on familiarity, are processed in an ‘anterior temporal system’ including but not limited to the perirhinal cortex, anterior hippocampus, and lateral entorhinal cortex. In contrast, the ‘posterior medial’ cortex including posterior hippocampus, parahippocampus, retrospenial cortex and angular gyrus is involved during recollection of contextual information associated with an item. However, it may be overly simplistic to completely segregate these systems, as neural reinstatement of item memories, such as faces, has been seen in posterior medial regions including the angular gyrus (Lee & Kuhl, 2016), meaning that anterior and posterior memory networks likely interact during item memory retrieval. However, the neurocognitive differences between item recognition versus elaborate episodic memory retrieval may lead to qualitatively different processes underlying the updating of these memories.

Differences between item and elaborate episodic memory updating may also be critical when considering the importance of neural reactivation to induce memory updating. For example, it has been shown that the amount of information remembered during retrieval correlates with increased neural reinstatement (Johnson et al., 2009; Leiker & Johnson, 2014). As mentioned, memory updating is critically related to the extent to which memories become reactivated (Gershman et al., 2013; Hupbach et al., 2008; Lee et al., 2018; Sederberg et al., 2011; St. Jacques et al., 2013), therefore the amount of neural reactivation could be limited during item compared to elaborate episodic memory retrieval simply because there is less information that participants can retrieve (e.g. Leiker & Johnson, 2014). Such a suggestion could be made when considering the retrieval processes that are likely to occur during recall of elaborate episodic memories vs. recognition of items. Specifically, it is likely that recalling the content of elaborate memories relies more on reconstructive recollection mechanisms, whilst recognising items will often engage familiarity-based judgements without context retrieval, especially if the context associated with items is the same across different memories (though see Thakral et al., 2015 for evidence of reinstatement during both familiarity and recollection). Thus, it is important to examine the behavioural and neurocognitive mechanisms of item memory updating, induced via recognition-based retrieval judgements, to understand how these types of memories become updated through repeated retrieval attempts.

1.5 Face recognition and face space

The present thesis used face stimuli to examine how item memories become updated through repeated recognition attempts. These face stimuli were computer generated images created using face space models of recognition. The face space model is a hypothetical system consisting of a multi-dimensional space that the human brain uses to represent faces (Valentine, 1991; Valentine, Lewis & Hills, 2015). These dimensions reflect normally distributed but non-specified characteristics, with faces having a unique location in space according to their position on each of the dimensions. Faces located proximally to each other are thought to be more perceptually similar, whereas faces that are at more distant locations are more perceptually dissimilar. Furthermore, it is thought that face space develops according to the type of faces that an individual perceives in their lifetime. Within an individual's face space, it is assumed that more ‘typical’ faces are located more densely at the centre of the face space distribution, whereas distinctive faces are located at more sparse locations away from the centre of face space (Valentine et al., 2015).

Additional assumptions of the face space models suggest that faces have a varying degree of error that surrounds the location of face representations in space (Valentine, 2001; Valentine et al., 2015). The size of representation errors is thought to relate to the quality of how faces are encoded initially. That is, if faces are encoded in optimal learning conditions (e.g. longer presentation times, viewing upright faces, frontal views), then the error size is reduced. In contrast, if faces are encoded in sub-optimal learning conditions (e.g. shorter presentation times, viewing inverted faces, side views), then the encoding error surrounding faces will increase. The size of representation errors is important as the overlap of face representations in space, more likely with larger representational errors, may lead to errors during face recognition attempts.

Face space models postulate that, during face recognition, a signal is generated from a match between faces seen during recognition and a stored representation in face space (Valentine, 1991). The strength of this recognition signal is then used as a basis of recognition judgements for individuals to indicate whether they have encountered a face previously. If a stored representation is located at the centre of space, where more facial identities are clustered, then faces with a larger representation errors will likely overlap with other face identities. This overlap may then lead to an increased difficulty in accurately recognising target faces, and may lead to falsely recognising non-target faces as those seen previously. Such an issue is less apparent for more distinctive faces, where a larger error will be less likely to overlap with other face identities, making recognition of distinctive faces easier (Valentine, 1991).

The face space assumptions described above have been formalised in Lewis' (2004) face-space-R model. This computational model was proposed to account for the recognition of familiar faces (e.g. celebrities, family members) and unfamiliar faces (e.g. previously unknown such as criminal suspects), and in particular how faces are transformed from unfamiliar to familiar identities. According to Lewis (2004), each face in space has a representational strength parameter. Faces that have been perceived a limited number of times (i.e. unfamiliar faces) have weaker representations whilst highly familiar faces have increased representational strengths. During recognition, each face is associated with a recognition strength parameter, with individuals endorsing a face as seen before if the recognition parameter meets a threshold value. Recognition is more likely to occur for faces that have stronger representation strengths in the model. Furthermore, a larger recognition parameter value is necessary to endorse typical faces as being seen previously due to the volume of face images that are present at the centre of face space. These theoretical issues raised by Lewis (2004) will be considered again in the general discussion chapter (Chapter 7) to provide an interpretation of experimental data from Chapters 2-6.

An important practical application of the face space model has been for the use of facial composite systems in eyewitness testimony, such as EFIT-V (Solomon, Gibson & Mist, 2013). These ‘holistic’ composite systems are designed using an ‘artificial’ face space constructed by performing a principal component analysis on a set of human face pictures. The artificial face space is represented by multiple independent components that each capture unique aspects of face variability from the input face set, from which new artificial images such as facial composites can be generated by assigning weights to each component (Solomon, Gibson & Mist, 2013). In creating facial composites, eyewitnesses begin with the presentation of a set of face images randomly generated from the artificial face space. Witnesses then select the faces they think best resemble the suspect, from which a further set of composites are generated (see Solomon, Gibson & Mist, 2013; Mist, Gibson & Solomon, 2015 for specific details). This iterative process continues until witnesses develop a composite they think best matches their memory of the suspect’s face. Research has shown that using composite systems such as EFIT-V can improve eyewitness memory of a suspect’s face during subsequent recognition attempts (Davis et al., 2014; 2015), emphasising the practical value of using artificial face space modelling in psychological research. An artificial face space model (informed from Mist et al., 2015) was adopted for the current thesis as a means to create face images, with details on face stimuli including a discussion on the benefits of using such stimuli in Chapter 2 (section 2.1).

1.6. Aims and hypotheses of the thesis

Using face images created from artificial face space models, this thesis aimed to investigate the neurocognitive mechanisms of face memory updating. As mentioned, one factor investigated whether subjective retrieval processes during initial face recognition attempts, measured using confidence judgements, modulated the updating of face memories. It was predicted that high confidence judgements during initial recognition would increase the likelihood that participants would select the same face during a subsequent recognition task (consistent with St. Jacques et al., 2013; St. Jacques & Schacter, 2013). Importantly, this effect was predicted when both target (i.e. correct recognition) or distractor faces (i.e. incorrect recognition) were selected during initial recognition, emphasising both the beneficial and detrimental impact of retrieval on face memory updating.

In a separate set of experiments, this thesis aimed to assess whether face memory updating is enhanced when memories are retrieved rather than re-studied. Consistent with the retrieval practice literature (Marsh et al., 2007; Roediger & Butler, 2011; Roediger & Marsh, 2005), it was predicted that memory updating would be increased when participants selected faces following retrieval attempts, compared to selecting faces that participants were told to encode or selecting faces according to a distinctive judgement. Finally, the neural correlates underlying face memory updating were investigated with ERP and time-frequency analysis of an EEG experiment similar to the confidence experiments described in the previous paragraph. It was hypothesised that, during initial recognition attempts, neural correlates of updating would be expected only if participants went on to select the same/similar faces at later recognition (e.g. Bridge & Paller, 2012; Liu et al., 2018). These neural correlates should differ to those corresponding to objective retrieval accuracy and subjective retrieval confidence, which were hypothesised to be reflect the reactivation of memories during retrieval (Bridge & Paller, 2012; Liu et al., 2018). Next, I will summarise the structure of this thesis before the first experimental chapter (Chapter 2).

1.7 Structure of the thesis

The present thesis aimed to investigate the mechanisms of face memory updating across 5 experimental chapters. Initially, Chapter 2 provides a methodological background of several key elements of the current thesis, including details of face stimuli used (Section 2.1). The second chapter also provides details of statistical analysis and presentation of results throughout the thesis, including a discussion on how face memory updating was operationalised (Section 2.2.1), a description of how results are presented using scatterplots (Section 2.2.2), a presentation and discussion of the use of Bayesian statistics for behavioural data analysis (Section 2.2.3) and a description and discussion of EEG data analysis used in Chapters 5 and 6 (Section 2.2.4).

Chapter 3 examines how objective and subjective retrieval processes interact to induce face memory updating. These experiments presented several target faces during a learning task, with a recognition task (target faces shown amongst four distractor faces) completed in two separate tests. During Test 2, recognition performance was analysed according to whether participants made the same response to those made in Test 1 (recognition bias), separately for Test 1 recognition responses that were correct with high or low confidence, or incorrect with high or low confidence. Both experiments in Chapter 3 aimed to assess whether increased recognition confidence during Test 1 led to larger test 2 recognition bias measures. The only difference between Chapter 3 experiments was how confidence was measured during recognition tests, with a categorical scale used in Experiment 3a, and a continuous scale used in Experiment 3b.

The aim of the experiments in Chapter 4 was to assess whether active retrieval of face memories is a critical factor to induce face memory updating. In all experiments from Chapter 4, participants completed two blocks of learning, refresh and final recognition tasks. Within these experiments, participants' performance during final recognition measured the extent to which participants correctly recognised target faces (recognition accuracy) or falsely recognised a distractor face during final recognition that was also selected during the preceding refresh task (recognition bias). Recognition accuracy and bias measures were compared separately for blocks one and two, allowing the comparison of these measures as a function of the task completed during refresh. In Experiment 4a, the refresh task in block one required participants to remember which face were targets (i.e. active retrieval), whereas the block two refresh task asked participants to encode one of five faces to-be-remembered for final recognition (i.e. passive re-study). In Experiment 4b, the refresh task in both blocks included the active retrieval task to examine whether order effects of block contributed to differences in final recognition measures. Finally, block one in Experiment 4c required participants to select one of five faces according to which face was most distinctive (referred to as "select refresh", involving refreshing memory through self-choice, rather than active retrieval or passive re-study), with block two refresh including the passive re-study task. Across all three experiments, the effect of refresh task (active retrieval, passive re-study, select refresh) on subsequent memory was compared to assess whether active retrieval of face memories led to increased face memory updating.

In Chapters 5 and 6, the thesis moves on to examine the neural correlates of face recognition with ERPs (Chapter 5) and oscillations (Chapter 6). The goal of these chapters was to examine the neurocognitive mechanisms underlying the objective, subjective and updating mechanisms during the recognition of faces. By using ERP and oscillatory analyses, the two chapters offer complementary yet unique insights underlying these recognition processes by using a similar experimental paradigm to that in Experiment 3b. In Chapter 7, the results from experiments in Chapters 2-6 are discussed in light of theoretical models and assumptions from both episodic memory and face recognition that have been discussed in this opening chapter.

Chapter 2 - Summary of methodology for behavioural and EEG experiments

2.1. Face stimuli

The experiments included in this thesis examined retrieval-induced updating of face memories with computer generated face stimuli produced with a method based on face space models of face recognition (Valentine et al., 2015). The face space is a hypothetical system that human brains use to represent faces, consisting of a multi-dimensional space where each dimension reflects normally distributed but unspecified facial characteristics. Within this space, individual faces have a unique location; more typical faces are clustered around the centre of the space, whilst distinctive faces are located away from the centre of the space (Valentine et al., 2015).

A practical use of the face space model is in the context of eyewitness testimony, specifically for face composite systems, such as E-FITV (Solomon, Gibson & Mist, 2013). These composite systems generate an artificial “image face space” by principal components analysis on images of real faces. The produced components each account for a certain amount of variance in the face image set. A selection of these components are then used to create novel, composite face images by attributing different weights to each component. The algorithm used for the present set of experiments allowed the production of such face composites, under the control of several parameters. First, the number of components used to create the face space can be varied so that only components that account for a meaningful portion of face variance are included in the model. Second, the distance that each generated face image is from the centre of the face space distribution can be controlled (referred to as the radius), as well as the multidimensional distances between each generated face image in directions orthogonal to the radius.

Using the image face space method, a number of ‘face sets’ were created for the current experiments. Each set contained a number of face images that were slight variations from one another. Face stimuli varied in feature morphology (type of features and spacing between features) but not by other properties such as face shape (which was fixed across images), skin tone (as face space was derived from Caucasian faces) or eye colour (see Figure 2.1). In creating face sets, a number of face locations were initially defined that were sampled at an equal distance from the centre of face space, in order to control to face image distinctiveness (i.e., faces located further from the centre are more distinctive, Loffler, Yourganov, Wilkinson & Wilson, 2005; Valentine et al., 2015). For each face location, a number of additional faces were created that were randomly sampled from a distribution surrounding each face location, but still retaining an equal distance from the centre (all faces were sampled from a “hypersphere” around the centre). For example, Figure 2.1 shows a face set where face “A” was defined as the original face location, with the four additional faces randomly sampled from within a limit of +/- 4 standard deviations from the face A location. The purpose of sampling within a limit based on standard deviations is to ensure that face images vary from the original to a certain maximal amount (if the limit is too large, the faces no longer resemble each other). Specific details of parameters used to create the face sets within each experiment can be found in the relevant experimental method section.

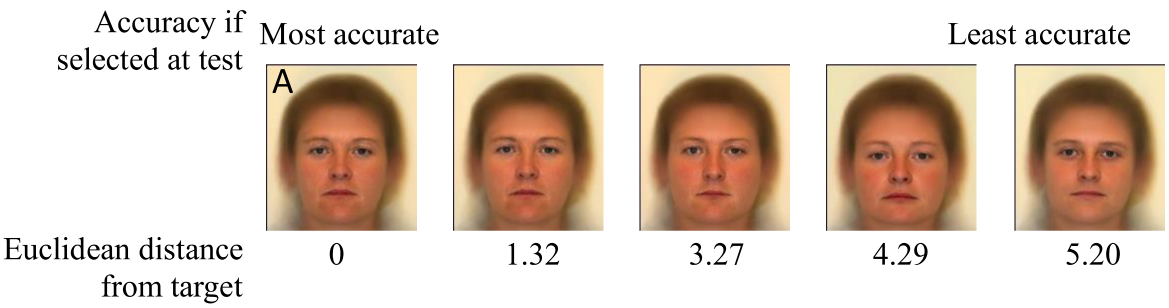

Figure 2.1. Example face set used in the present thesis. Face A, designated as a target face, was created along with four distractor faces that varied in perceptual dissimilarity, which is related to a larger Euclidean distance in image face space. N.B. for all experiments, face stimuli were presented at a visual angle of 5.12 x 5.88 when participants sat 75cm from the screen.

One key motive for using stimuli generated by this face space method is to allow the exploration of a continuous measure of memory processing. For example, as shown in Figure 2.1, face sets show a degree of variability in perceptually dissimilarity. Greater dissimilarity between two faces can be represented as a larger distance in multidimensional face space, which can been quantified by measuring the Euclidean distance between the face locations (Tredoux, 2002). For the current set of experiments, one face within a set was a ‘target face’ that is shown during learning and recognition tasks. The remaining faces within a set were ‘distractor faces’ that were only shown during recognition tasks. During these recognition tasks, participants’ performance was quantified by the Euclidean distances between a pair of images as measures of degrees of recognition accuracy and recognition bias. Recognition error (where smaller values indicate more accurate recognition) was measured as the Euclidean distance between the target face within a set and the face participants selected as their recognition response. For example, as demonstrated in Figure 2.1, if participants correctly recognised the target face, recognition error would be measured as 0. Conversely, if a distractor face was selected at recognition, then recognition error would be the distance between the target and the chosen distractor image (e.g., Euclidean distances of 1.32 – 5.20). That is, the more dissimilar a selected face is to the target face, the greater recognition error is with the Euclidean measure. Consequently, the Euclidean distance measure assesses the extent to which recognition memory is accurate or erroneous, as opposed to categorical measures such as proportion correct responses that indicate whether responses are correct or incorrect regardless of the magnitude of errors. Thus, Euclidean distance provided a continuous measure of retrieval accuracy in the current research, analogous to continuous measures that have been used in studies examining the cognitive (Harlow & Donaldson, 2013; Harlow & Yonelinas, 2016) and neurocognitive mechanisms (Murray et al., 2015; Richter et al., 2016) underlying recollection precision.

To summarise, the key motives for using artificial images from face space models allowed for controlling face image variability, in addition to assessing face recognition performance with continuous metrics. Given that only few studies have used these face images in the context of episodic memory research prior to this thesis, several pilot experiments were conducted. Firstly, the learning procedure used within the reported experiments was validated to ensure faces were sufficiently encoded (see Appendix A). In addition, validation experiments were conducted to validate the assumption that a larger Euclidean distance between face images corresponds to a larger perceptual dissimilarity.

2.2. Statistical presentation and analysis

For both behavioural and neuroimaging analyses, custom Matlab scripts (version 17a) were used for data extraction. Statistical analyses was completed with SPSS 25 or

JASP (version 0.10). Behavioural data were analysed with GLM models (ANOVAs, t-tests, correlations, etc.) and using frequentist statistics for significance testing, which is standard in Psychology and Neuroscience so does not require further explanation. However other methods used during data analysis were more uncommon or customised for the current project, such as the use of Bayesian statistics, and various EEG analysis techniques and parameters. These less common methods are outlined and justified in the following sections.

2.2.1. Behavioural analysis of face memory updating

In order to measure face memory updating, a ‘bias’ measure can be calculated from participants’ performance across multiple recognition tasks (e.g. Bridge & Paller, 2012; Bridge & Voss, 2014). For example, in the repeated recognition paradigm presented throughout the current thesis, participants were required to select one of five faces they recognised during ‘Test 1’. The same recognition task was ‘repeated’ for some of the trials in Test 2[3], meaning that participants provided two recognition responses for these trials across two tests. For a given trial, if participants selected the same face in both tests, this would be considered a ‘biased’ response in the sense that participants are ‘biased’ during Test 2 towards recognising the same face that was selected during Test 1. Across all repeated recognition trials, a ‘proportion bias’ score can then calculated for each participant. For the Experiments 3a and 3b, proportion bias scores were calculated separately for trials when recognition responses in Test 1 were correct (made with high vs. low confidence) or incorrect (made with high vs. low confidence).

---

[3] For all experiments 2/3rds of trials were assigned to the repeated condition. So for experiments in Chapters 3 and 4 (where 30 face sets in one block), 20 face sets were ‘repeated’ and 10 sets were not. These ratio’s were chosen as an optimal balance for ensuring enough trials were assigned to the repeated and baseline conditions.

For Experiments 4a, b and c in Chapter 4, face memory updating was compared during a final recognition task according to the type of task completed in a previous ‘refresh’ phase (see Sections 4.1.1, 4.2.1 & 4.3.1 for Method details in Experiments 4a, b & c respectively). The refresh task, shown after an initial learning phase, presented one target face with four distractor face images. Participants were required to remember which face was shown during learning, or were asked to encode one of the five faces to-be-remembered for the final recognition task. The effect that these refresh tasks had on memory updating was assessed using a final recognition bias measure, with the prediction that asking participants to remember faces during refresh would enhance final recognition bias compared to re-studying faces during refresh, under the assumption that retrieval rather than re-study is a more powerful learning mechanism to induce episodic memory updating (e.g. Roediger & Butler, 2011). However, the final recognition bias score used in Experiments 3a and b was not used for Experiments 4a-c, as this measure may have been confounded by the type of refresh task. That is, for Experiments 4a and 4c, a ‘passive’ re-study task was designed that told participants to encode one of five faces, with this face determined by the responses made in a refresh task in a previous block where participants were free to select any of the five faces within a trial (based on memory during active retrieval, or a distinctiveness judgement during selection refresh). Within passive re-study trials, it is possible that participants may have recognised one of the faces as the target, but then be told to encode a different face that they did not recognise, resulting in a conflict between their incidental recognition of the target and what they are being told to remember for a subsequent test. In this situation, participants could have decided to ignore the distractor face they were told to encode and instead focus on the face they recognised from learning. Should this occur, a lower final recognition bias score following passive re-study would not result from differences in processes engaged during retrieval/re-study, but instead be due to participants ignoring the cued distractor faces during passive re-study. In contrast, the proportion bias score was used for the Experiment 3 analyses because this bias measure would be less likely to be confounded between the critical conditions in these experiments (i.e. between high and low confidence conditions). That is, for both high and low confidence judgments, recognition judgements were based on genuine retrieval attempts, meaning that a similar process (i.e. retrieval attempts) contributed to the selection of faces during initial recognition for both high and low confidence judgements).

In order to avoid the confounding influence of refresh task on proportion final recognition bias, a modified bias score was created for behavioural data analysis in Experiments 4a, b and c. These bias measures were calculated only on refresh trials where participants selected a distractor face. For these “prior error” trials, a bias score calculated the proportion of trials where participants selected the same distractor face during final recognition. Additionally, a non-bias score calculated the proportion of prior error trials where participants selected a different distractor face during final recognition to the one selected from refresh. The averaged non-bias score (dividing the non-bias score by the number of possible non-bias options) was deducted from the bias score providing a bias difference measure. More positive bias difference measures indicate participants were more likely to repeatedly select the same distractor, compared to switching recognition responses to a different distractor face, during final recognition. During final recognition, it could be expected that, relative to all of the distractors faces shown during refresh, participants would more likely pick the face they were told to encode (indicated by a positive bias difference score), providing some evidence that cued faces were not simply ignored during passive re-study.

The bias difference measures were only calculated for “prior error” trials (rather than all refresh trials or “prior correct” trials) as if both correct and incorrect refresh trials were analysed, these bias measures would confuse two different forms of episodic memory updating; when target faces were repeatedly correctly recognised compared to when distractor faces updated memories. Thus, by analysing bias for “prior error” trials only, this measure focused on the updating of memories purely from repeated incorrect recognition attempts. For “prior error” trials during refresh, the extent to which participants corrected these errors during final recognition (i.e., participants switched responses to select a target during final recognition), was not analysed again due to a confound of refresh task. During active retrieval, selecting a distractor face image was based on a genuine yet erroneous retrieval attempt, suggesting that participants had an imperfect memory of target memories for these trials. However, and similar to the issues raised previously, the “prior error” trials for re-study conditions may contain a mixture of trials where participants could or could not have remembered the target face if these trials were in the active retrieval condition. In this sense, if the proportion of final recognition trials where target faces were selected was calculated for “prior error” trials, then target selection rates would be expected to higher following the re-study tasks. These results would be interpretable with a retrieval-induced updating hypothesis, as it could be predicted that participants would be less likely to select a target following active retrieval, if their memory had been updated by distractor faces that were falsely recognised during refresh. However, these findings would not be due to the refresh task manipulation but instead be due to having more strong, durable target face representations for trials in re-study vs. retrieval conditions. Overall, these arguments justify the use of the final recognition bias difference score for “prior error” trials in the Experiments of Chapter 4.

Whilst the bias measures used in Experiments 3a-b and 4a-c were used to operationalise memory updating, these measures do not reflect one particular mechanism of memory updating over another. For example, face memory updating could be achieved by one of several mechanisms; a face memory representation could be directly distorted (via a reconsolidation like process). Alternatively, updating could occur via interference of competing memory representations (between faces encoded during learning vs faces encoding during test 1). Thus, bias could be achieved by one of many underlying processes, which justified the EEG experiment presented in Chapters 5 and 6 to understand the neurocognitive processes underlying face memory updating. Furthermore, and regardless of the mechanistic process underlying face memory updating, the bias measures reported in these experiments provided a useful metric to assess the extent to which participants were biased during repeated recognition attempts.

2.2.2. Behavioural data plotting

Within all experimental chapters, behavioural data related to face recognition accuracy/updating are presented as scatter plots. These plots show the mean value for conditions (entered into statistical analysis), with scatterplots presented on top to demonstrate the distribution of participant-level scores within conditions. These plots used the plotSpread function in Matlab. In addition, 95% high density intervals (HDI) estimated with a Bayesian bootstrap were presented for each condition (calculated using the Matlab RST toolbox, RST Toolbox Team, 2015). The 95% HDI shows the 95% probability of the variable parameter lying within the upper and lower limits on the box plotted. The advantage of using these plots are two-fold. First, these plots show not only the mean value of a condition, but also the distribution of participant data points to demonstrate the variability in participant performance within a certain condition. Second, the use of a 95% HDI is considered by some as a more intuitive estimation of the population parameters (Morey, Hoekstra, Rouder, Lee & Wagenmakers, 2016). That is, 95% confidence intervals states that, if the same study was repeated several times in the future, 95% of these studies will contain the population parameter. However, and possibly more in-line with what researchers infer from a confidence interval, a 95% HDI states that there is 95% probability that a single data set contains the population parameter (Morey et al., 2016). The latter appears more intuitive and in-line with the aims and interpretations of the present research, and thus was used in the subsequent data plotting.

2.2.3. Behavioural analysis with Bayesian statistics

For experimental Chapters 3, 4 and 5, Bayesian statistical analysis was performed as an addition to frequentist analysis of behavioural data performance. Bayesian statistical inference determines the extent to which data supports a pre-determined hypothesis that is specified as a statistical model. Crucially, Bayesian analysis, unlike frequentist analysis, considers at least two competing hypotheses: most typically a hypothesis defining a difference between conditions (i.e., alternative hypothesis), and a hypothesis defining no difference between conditions (i.e., null hypothesis).

One form of Bayesian analysis, the Bayes Factor (BF), calculates the ratio of evidence in favour of the alternative hypothesis (H1) compared to the null hypothesis (H0), as opposed to traditional ‘frequentist’ statistics, such as null hypothesis significance testing (NHST) that determines the probability of the observed data occurring if the null hypothesis was true (see Dienes, 2011, for frequentist-Bayesian comparisons). Briefly, Bayes factors are calculated as:

$$\frac{P(H1|D) = P(D|H1)*P(H1)}{\underbrace{P(H0|D)}_{\text{Posterior}} = \underbrace{P(D|H0)}_{\text{Bayes Factor}}*\underbrace{P(H1)}_{\text{Prior}}}$$

where the numerator refers to the alternative hypothesis, and the denominator refers to the null hypothesis (adapted from Dienes, 2016). In this equation, the ‘posterior’ specifies the ratio of the probability of the alternative hypothesis being true versus the probability of the null hypothesis being true, given the observed data (D). The ‘prior’ specifies the ratio of the prior probability of the alternative hypothesis versus the prior probability of the null hypothesis. More specifically, priors allow a researcher to state the probability of a hypothesis occurring *prior* to data collection. The incorporation of prior information of the two hypotheses is considered a strength of Bayesian analysis, allowing researchers to fine-tune statistical analysis to incorporate expected/predicted effects prior to data collection. Finally, the Bayes factor specifies the extent to which evidence from observed data favours the alternative hypothesis versus the null hypothesis. In this equation, Bayes factors can range from 0 (evidence favours the null hypothesis) to 1 (evidence favours neither hypothesis) to ∞ (evidence favours the alternative hypothesis). According to Jeffreys (1961), a Bayes factor greater than 3 (or less than 0.3) provides convincing evidence in favour of the alternative (or null hypothesis), respectively (see Table 2.1 for full range of labels to interpret Bayes Factors).

Table 2.1. Labels for interpreting Bayes Factors.

Bayes Factor	Inverse Bayes Factor	Label
1-3	1-.33	Anecdotal
3-10	0.33-.10	Substantial
10- 30	0.10-.03	Strong
30-100	0.03-.001	Very Strong
100>	<0.001	Decisive

Note. Table adapted from Jarosz and Wiley (2014).

Recent software developments have enabled greater access to compute Bayes factors for a range of statistical tests (including ANOVA, t-test and correlation), such as the statistical software package JASP (JASP Team, 2018). With JASP, the only requirements to compute a Bayes factor are variables of interest, and to specify the 'prior' distribution of the H1 and H0. The prior distribution can be specified by defining three components. Firstly, the direction of the expected effect (variable 1 greater than variable 2, variable 1 equal to variable 2, or variable 1 less than variable 2) can be specified. Second, the form of the Bayes factor can be denoted as BF_{10} (larger numbers provide evidence in favour of the alternative hypothesis) or as BF_{01} (larger numbers provide evidence in favour of the null hypothesis). Finally, and arguably the most difficult decision to make, is defining the prior distribution for the alternative hypothesis (the prior for the null hypothesis is fixed for all Bayes factor tests in JASP).

A variety of options are given in JASP in order to define the prior distributions, with prior specifications dependent on the type of statistical analysis. For ANOVA analyses, three prior distributions are specified; r scale fixed effects (i.e., between-group variance), r scale random effects (i.e. within-group variance) and r scale covariates (mixture of between and within-group variance). Varying such parameters with ANOVA has little meaningful change on the subsequent Bayes factor (Rouder, Morey, Verhagen, Swagman, & Wagenmakers, 2016), thus for Bayesian ANOVA's in the present thesis the default options are not changed. In contrast, for Bayesian analysis of a t-test, the prior is specified by the predicted effect size between two conditions, based on previous analyses or theoretical reasoning. A default option is provided by JASP (Wagenmakers et al., 2018), which defines the prior distribution as a Cauchy distribution (recommended by Rouder, Speckman, Sun, Morey & Iverson, 2009), centred at an effect size (δ) of 0 with a scaling factor of 0.707 (meaning that 50% probability is assigned to effect sizes within 0.707 standard deviations from 0; see Figure 2.2). However, the use of a default prior distribution negates a critical advantage of Bayes factors; that being the ability to use subjective, theory-driven predictions for a particular research question. Fortunately, JASP allows one to define 'informed' prior distributions, i.e., prior distributions that can vary according to the specific predictions of an individual research question.

An informed prior distribution in JASP can define the distribution as a Cauchy, normal or t distribution. For each distribution, the central location and width (a.k.a scaling factor) of the distribution are defined as the expected effect size difference. For the present thesis, Cauchy distributions were centered at the effect size found for the corresponding analysis in previous experiments (using Hedges *g* effect sizes[5]). The scaling factor of these Cauchy distributions was the default prior width of 0.707 in JASP. The specific prior

---

[5] Hedges *g* effect sizes were used instead of typical effect measures (e.g. Cohen's d) as an unbiased effect size measure (Lakens, 2013)

distributions are reported with each test and are denoted as $B_{10(0.5,0.707)}$, with B_{10} indicating that the Bayes factor reported evidence in favour H1 over H0, and $_{(0.5,0.707)}$ stating that the prior Cauchy distribution was centred at an effect size of 0.5, with a scaling factor of 0.707. Furthermore, the direction of each test (one or two-tailed) are denoted as B_{+0} or B_{10}, respectively, according to the specific predictions for each t-test.

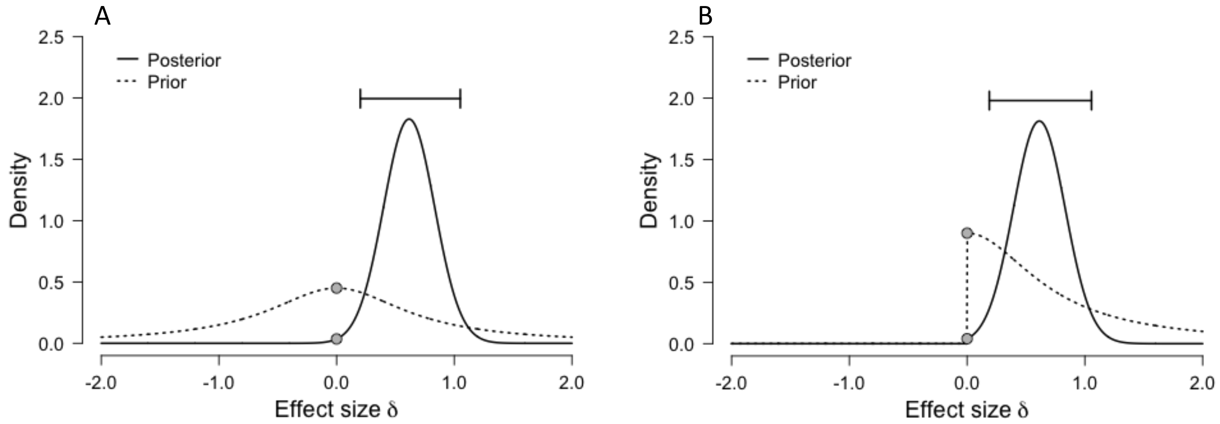

Figure 2.2. Plot of prior and posterior distributions produced in JASP. The prior distribution (dotted line) specifies an alternative hypothesis that is predicted to occur in future data, based on theoretical prediction or effect sizes from previous studies. The prior distribution can be defined by a two-tailed (Panel A) or one-tailed (Panel B) to specify the direction of expected effect size differences.

In summary, Bayes Factors were calculated for each ANOVA and t-test analysing behavioural performance. This was applied in Experiment 3b, Experiments 4b and 4c, and Experiment 5. Bayesian analysis was not used in Experiments 3a and 4a, owing to the fact that these experiments were the first conducted within their respective chapters, thus an informed prior could not be justified for these experiments. Whilst Bayes Factors could have been calculated using default priors in Experiments 3a and 4a, or from effect sizes of the most relevant published research, neither of these options were justified on the basis that, given the novelty of experiments and stimuli in this thesis, these effect sizes may have been substantially different to the true population effect size relevant for the present data. However, for all other experiments, an informed prior could be specified based on the effect size found for statistical tests based on results from Experiments 3a and 4a[7]. For each t-test, the ‘prior’ effect size specified the centre of the prior distribution, with the prior width set to the default of 0.707.

2.2.4. EEG analysis

For the EEG experiment conducted for the present thesis, ERP and oscillatory effects were analysed and are presented separately in Chapter 5 and 6, respectively. These two analytical strategies were used to offer complementary approaches to the analysis of EEG data.

EEG pre-processing. From raw EEG recordings, several pre-processing steps were conducted with the EEGLAB toolbox (Delorme & Makeig, 2004), in order to prepare and clean EEG data for later analysing ERPs and oscillations. The critical steps taken for EEG cleaning in the present thesis were: filtering data, removing noisy EEG with independent component analysis (ICA), and interpolating noisy electrodes. Filtering EEG data is necessary to remove very low (with high-pass filter) and very high (with low-pass filter) frequencies from the EEG data that are likely to be “noise” artefacts (not generated by brain activity). For the present thesis, filtering in EEGLAB used a Hanning windowed finite impulse response filter. With ICA, the recommended option from EEGLAB was used, this being the runica function. This function uses an infomax ICA algorithm (Bell &

---

[7] It is important to acknowledge that the effect sizes from Experiments 3a and 4a may not be a true reflection of the population effect size. However, with all Bayes Factor calculations, robustness plots created in JASP were used to verify the stability of Bayes Factors across a range of prior distributions so as to not base any conclusions from a single effect size.

Sejnowski, 1995) to decompose EEG into spatially filtered components that reflect unique contributions to the recorded EEG signal. The primary goal of ICA for pre-processing is to identify and remove several noise components from the EEG, specifically components corresponding to eye blinks, eye movements and high-frequency noise (from muscle activity or faulty electrodes). Finally, electrode interpolation was used to replace EEG recorded from faulty electrodes, with interpolated electrodes also included for statistical analysis. No more than 10% (six electrodes) were interpolated for a single EEG session, in order ensure that EEG recordings were not dominated by artificial EEG signals.

Event-related potentials. Event-related potentials (ERPs) reflect the electrophysiological voltage of neural activity over time (Kappenman & Luck, 2012). ERPs recorded from the scalp show either positive or negative deflections that reflect an on-going neurocognitive process that is time-locked to stimulus or response onset. ERPs are computed by averaging together segments of EEG that are time-locked to a particular event, such as a stimulus or response (e.g. correct/incorrect recognition, high/low confidence). ERPs can then be compared between conditions to identify the temporal dynamics of neurocognitive processes underlying behavioural performance, such as recognition and updating of face memories. For the present thesis, EEGLAB (Delorme & Makeig, 2004) calculated and extracted ERPs from collected EEG data.

The ERP technique was adopted in Chapter 5 as this method has been the popular technique since the 1960’s as an electrophysiological analysis approach (Luck, 2005). Plenty of studies since this time to the present have applied ERPs to studies of episodic memory (Wilding & Ranganath, 2011), thus a rich literature can be referenced to help interpret the neurocognitive processes related to retrieval-induced updating of face memories in Chapter 5. One particular advantage of the ERP technique, specifically compared to alternative electrophysiological analysis such as time-frequency analysis, is the increase in signal-to-noise ratio improvement that the ERP provides by averaging the time-locked EEG activity. With sufficient trials, the ERP thus provides a more reliable measure of neural processes corresponding to cognitive operations during retrieval and updating of face memories.

The signal-to-noise ratio of ERP data is an important feature to consider for such analyses, in particular to ensure that experiments have sufficient sample size, and each participant has an adequate number of trials per condition, to maximise the signal-to-noise ratio. With regards to trial numbers per condition, the more trials a condition has, the higher the signal-to-noise ratio. Over recent years, ERP literature has debated the minimum number of trials needed to find a reliable ERP effect. A recent study by Boudewyn et al. (2018) argues that signal-to-noise ratios depend on a mixture of factors including sample size, trial numbers per condition, and the effect size of ERP differences. Whilst this paper focused on only two specific ERP components (the error-related negativity and the lateralised readiness potential), the results imply that the power to detect a difference between ERP conditions reduces substantially with smaller effect sizes (1-3μV between conditions) and lower trial numbers per condition (8 trials). These implications are considered in light of the present ERP experiment, with a discussion on signal-to-noise ratio implications provided in the discussion section of Chapter 5 (see section 5.3).

Oscillations. Time-frequency analysis assesses the pattern of oscillatory activity present in EEG. EEG oscillations are rhythmic patterns of neural activity generated by populations of neural ensembles that can be measured from scalp electrodes (Bastiaansen, Mazaheri & Jensen, 2011). Neurons within an ensemble can either synchronise (neurons fire simultaneously) or desynchronise (neurons reduce in firing rates of do not fire simultaneously). Oscillations are characterised by their frequency of cycles per second and are often divided into specific frequency bands as activity changes tend to dissociate between these frequency bands and correlate within bands. Therefore, these bands are thought to have different functional characteristics. The common bands used are: delta (1-4Hz), theta (4-8Hz), alpha (8-12Hz), beta (12-30Hz) and gamma (30-90Hz). The present thesis focused on analysing theta, alpha and beta oscillations. Slow delta oscillations were not possible to extract (due to constraints of EEG decomposition, see below), and furthermore, delta oscillations are likely to contribute to the ERP (Bastiaansen, Mazaheri & Jensen, 2011), thus extracting slow-wave oscillations was not considered a priority since they may not add much novel information to that shown in ERPs. Fast-wave oscillations in the gamma-band, despite having strong links to episodic memory processing (Axmacher et al., 2010; Friese et al., 2013; Köster et al., 2014; Lisman, 2010; Nyhus & Curran, 2010; Parish et al., 2018; Staudigl & Hanslmayr, 2013), were not analysed as the EEG recording in the present experiment was filtered from 40Hz and above, thus frequencies in these higher bands, corresponding to the gamma frequency, could not be analysed.

Oscillations can either be evoked or induced with regards to stimulus/response onset. Evoked oscillatory activity is the oscillatory activity that is ‘phase-locked’ to a stimulus, whereas induced activity is not phase-locked to a stimulus (David, Kilner & Friston, 2006). For example, evoked oscillatory effects will be elicited by cognitive operations such as perception of a stimulus on-screen, and will typically be found shortly after stimulus onset. Conversely, induced EEG oscillations are generated by cognitive processes that are not strictly time-locked to stimulus, such as later stages of stimulus processing that may be more variable in onset (for example retrieval of episodic information associated with a stimulus that has been previously seen). Such retrieval processes may vary in temporal onset, depending on factors such as ease of retrieval, or the cumulative time of earlier processing stages. Crucially, given the varied temporal onset of induced oscillations, these effects will be averaged out during computation of ERPs

(Bastiaansen et al., 2011). On the other hand, evoked oscillations that are consistent in temporal onset will be present in ERP data (David et al., 2006), therefore ERPs and oscillatory correlates may provide unique insights into the neurocognitive processes related to the exact same task/data set.

Decomposing EEG into underlying frequency bands is important due to the differences in functional properties between evoked vs. induced neural mechanisms. It has been suggested that evoked and induced mechanisms reflect bottom-up vs top-down cognitive mechanisms, respectively (Chen et al., 2012; Tallon-Baudry & Bertrand, 1999). That is, bottom-up evoked processes may reflect stimulus-driven modulations that ‘feed forward’ from lower sensory (e.g. visual cortex) to higher cognitive processes (e.g. pre-frontal, parietal or temporal lobes). Conversely, top-down mechanisms from higher-order brain regions could modulate activity in lower-order regions in a ‘feed backward’ mechanism. These induced modulations may also represent oscillatory mechanisms that allow coupling (i.e., synchronisation) and uncoupling (i.e., desynchronisation) of brain networks underlying the retrieval and updating of face memories (Bastiaansen, Mazaheri & Jensen, 2011). In the context of episodic memory retrieval and updating, such bottom-up evoked and top-down induced modulations may reflect functionally distinct cognitive processes, as such a further discussion of these mechanisms will be provided in the discussion section in reference to the results from ERP (Chapter 5) and time-frequency analyses (Chapter 6).

For time-frequency analyses, on each EEG epoch, time-frequency decomposition was performed using morlet wavelets (Roach & Mathalon, 2008), a common choice of time-frequency decomposition in episodic memory literature. Morlet wavelets, constructed by multiplying a sine wave to a Gaussian function, act as a sliding time window across the epoch. Morlet wavelets are comprised of multiple wavelets that vary in the frequency to which the wavelet is applied, allowing a whole spectrum of frequencies to be estimated across the time window. The width of the wavelet, specified in FieldTrip, is consistent for all wavelet frequencies. This property means that the time length of wavelets for different frequencies will vary. For example, to estimate lower frequencies such as 4Hz, with a three cycle wavelet, the length of the wavelet will be 750ms (1000ms / 4Hz = 250ms x 3 cycles = 750ms), whereas the wavelet length for higher frequencies (such as 30Hz) will be 100ms (1000ms / 30Hz = 33.33ms x 3 cycles = 99.99ms). Consequently, wavelets with smaller widths have a finer temporal resolution at the cost of a poorer frequency resolution.

The choice of wavelet width for the present experiments was justified at three cycles in order to provide a reliable estimation of lower frequencies in the time-frequency decomposition. For example, a three-cycle wavelet at 4Hz has a time length of 750ms. If a wavelet had seven cycles (the default wavelet width in FieldTrip) then the length of this wavelet at 4Hz will equal 1,750ms, which spans more than half of the epoch length used in the present EEG experiment. Consequently, due to the length of the epochs in the present EEG experiment, a three-cycle wavelet width was necessary to decompose EEG into a time-frequency representation. Following time-frequency decomposition, grand averages of power values were calculated for each condition, separately for theta, alpha and beta bands. Conditions of interest were then analysed separately for each frequency band.

A relevant point regarding time-frequency decomposition considered the temporal length of epochs used in analysing time-frequency representations. The original epoch length measured 0.5s prior to face onset and lasted until 1.6s after face onset. These epochs encompassed the time that each face was on screen, as well as the fixation crosses shown prior to and after each face. This epoch length, however, is too short to estimate lower frequencies of interest, primarily those in the theta band. Given the importance of theta band in episodic memory processing (Nyhus & Curran, 2010), it was necessary for the pre-stimulus time period for each epoch to be extended in order for theta frequencies to be reliably decomposed. To allow for this, the pre-stimulus time period was extended from -0.5s to -1.2s using a ‘flipping’ procedure (see Chapter 6 for specific details). As noted in Figure 2.3, the pre-stimulus baseline was ‘flipped’, therefore only neural activity present during perception of a fixation cross was included in the extended baseline periods.

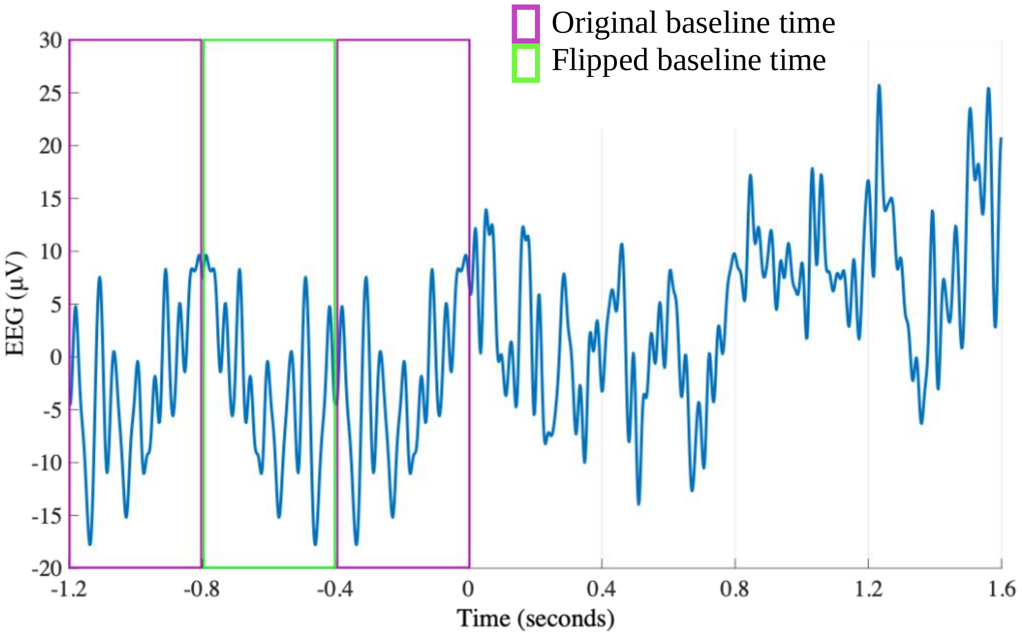

Figure 2.3. Example trial demonstrating the epoch ‘flipping’ procedure. For each epoch, the original -0.4s pre-stimulus baseline (highlighted in purple) was copied, flipped, and then appended to the beginning of the epoch (highlighted in green). The original pre-stimulus baseline was then appended to the start of the flipped baseline, resulting in time windows of -1.2s to 1.6s for each trial epoch within condition and participant.

Cluster-based permutation analysis of ERP and time-frequency data. Cluster-based permutation testing was used as the main statistical analysis method of both ERP and oscillatory data in the present thesis. Cluster-based permutation tests were used instead of traditional ANOVA analysis of EEG data for multiple time-windows and electrodes for several reasons. Firstly, cluster-based permutation tests overcome the multiple comparison problem of null-hypothesis significance testing with ANOVA. For a standard ERP analysis, a *p*-value can be calculated for a statistical test, such as a t-test, at all channels across the time window of the epoch. The same goes for analysis of time-frequency, with the addition of a frequency dimension in the analysis. Even if analysis is narrowed to *a priori* location, time or frequency points, the family-wise error rate of these t-tests is severely inflated with null-hypothesis significance testing, resulting in an increased likelihood of falsely concluding that a meaningful effect from EEG data when one is not present (i.e. inflated type I error). Bonferroni corrections could be applied to the alpha-threshold, however the number of t-tests run would result in a severely conservative bonferroni correction, meaning that meaningful effects in the EEG data could be disregarded (i.e. inflated type II error).

Cluster-based permutation testing overcomes the multiple comparison problem, and thus against inflating the type I error rate, in two critical ways (Maris & Oostenveld, 2007). Firstly, the cluster-correction approach is advantageous as this correction acknowledges that neighbouring spatial and temporal data points in EEG may be correlated (Sassenhagen & Draschkow, 2018). Spatially, true EEG activity at one electrode will be correlated with surrounding electrode neighbours as the same neural generator of signal in the brain propogates to several scalp locations (Sassenhagen & Draschkow, 2018). Therefore, EEG data from single electrodes are not truly independent, and may share common electrophysiological activity. In this regard, it makes sense to cluster together electrodes that share a common electrophysiological pattern, rather than treat them as independent data points, thus avoiding repetitive analysis of electrodes and time windows that may show the exact same effect to others.

A second benefit of the cluster-corrected permutation approach is the nature of the permutation statistical test. The permutation test, a non-parametric statistical test, has been shown to control the inflation of type I and II error rates with electrophysiological data (Pernet, Latinus, Nichols, Rousselet, 2015). One reason is the permutation test is not limited to assumptions of the data that parametric statistics are, such as independence of data, and the normality of distributions that samples are drawn from (assumptions which are more likely violated with EEG data). For these key reasons, cluster-based permutation tests were justified for the principal statistical test for ERP and oscillatory data analysis in the current thesis.

The second strength for using cluster-based permutation tests for the present EEG chapters refers again to *a priori* analysis. The present experimental paradigms were a novel approach to assess the neural correlates of face recognition and updating. Whereas previous studies have examined the ERP correlates of retrieval-induced updating (Bridge & Paller, 2012; Liu et al., 2018), it was unclear whether the same ERP effects found in these studies would be shown for the current experimental paradigm. Given that ERP correlates of face recognition differ to ERP correlates of name/word recognition (MacKenzie et al., 2018; Mackenzie & Donaldson, 2009), it was difficult to make specific predictions of what ERP effects would be related to retrieval-induced updating of face memories. Furthermore, at present, little-to-no research has used time-frequency analysis to measure neural mechanisms of retrieval-induced updating, though oscillatory correlates of general episodic memory retrieval were used to help interpret findings from Chapter 6. Consequently, cluster-based permutation tests were used as a data-driven analytical approach to the analysis of the ERP and oscillatory correlates of retrieval-induced updating, whilst controlling for type I errors.

The procedure for cluster-based permutation tests begins with t-statistics being computed between two conditions of interest at each spatio-temporal (ERP) or spatio-temporal-frequency data point (time-frequency). The t-values that exceed a specified alpha threshold (e.g., .05) are then selected for ‘clustering’, whereby neighbouring t-statistics within spatio-temporal or spatio-temporal-frequency proximity are combined into a single cluster of t-values. To be included in a cluster, each EEG channel (electrode) is required to have at least two neighbouring channels that also met the alpha threshold. By ensuring that clusters are formed from adjacent channels, this avoids spatially separated effects (possibly from distinct neural generators) being considered as part of one larger cluster.

For each defined cluster, a cluster-level statistic is computed, such as the maximum sum of t-statistics within the cluster. Cluster-level t-statistics can either be positive (condition one is larger than condition two) or negative (condition one is smaller than condition two). The cluster-level statistic then undergoes significance testing by creating a permutation null distribution using the Monte Carlo method. The permutation null distribution is created by collating all trials from the two analysed conditions. Two sub-lists are then made by randomly sampling from all trials, called a ‘random partition’, with a permuted cluster-level statistic calculated for these two lists. This procedure is then repeated a large number of times to create a null distribution of cluster-level statistics as estimated from each random partition. The originally observed cluster-level statistic is then placed within this distribution, and the proportion of permutation statistics that are more extreme than the observed cluster-level statistic is used as the permutation *p*-value.

If the permuted *p*-value is lower than a defined cluster alpha threshold (such as .05), then the null hypothesis that the two conditions are sampled from the same distribution can be rejected. Importantly, the spatial and temporal interpretation of the cluster requires an approximation based on descriptive data, as the significance of the cluster does not refer to specific time windows or locations of interest (Sassenhagen & Draschkow, 2018). Indeed, Sassenhagen and Draschkow (2015) suggest that the temporal onset of clusters may be underestimated, thus interpretations of specific cluster timings would be inappropriate. The descriptive nature of location, timing and frequency of significant clusters can be interpreted using ‘cluster-plots’ (see Figure 2.4). For the present thesis, cluster plots are presented as topographical maps for the analysed time period that show the scalp location and timing of significant clusters (indicated with yellow dots, see Figure 2.4). These plots only show clusters that last at least 100ms, meaning that for some analyses, a significant cluster lasting for fewer than 100ms is not plotted. Cluster plots can be produced for time-frequency data by averaging over a specified range of frequency values (e.g. averaging over frequencies within 4-8Hz to interpret a ‘theta’ cluster plot).

To further interpret the cluster-based permutation results, plots of descriptive statistics were produced for ERP (line plots) and oscillatory data (time-frequency representations and line plots). Line plots show the ERP (μV) or oscillatory power (dB) for each condition across the analysed time period at specific channels. Time-frequency representation plots, in contrast, plotted the power difference (dB) between two conditions across all frequencies (e.g. from 4-30Hz) across the analysed time period and averaged across all channels. These plots were produced to aid interpretation of power differences between conditions, visualised across frequency and time.

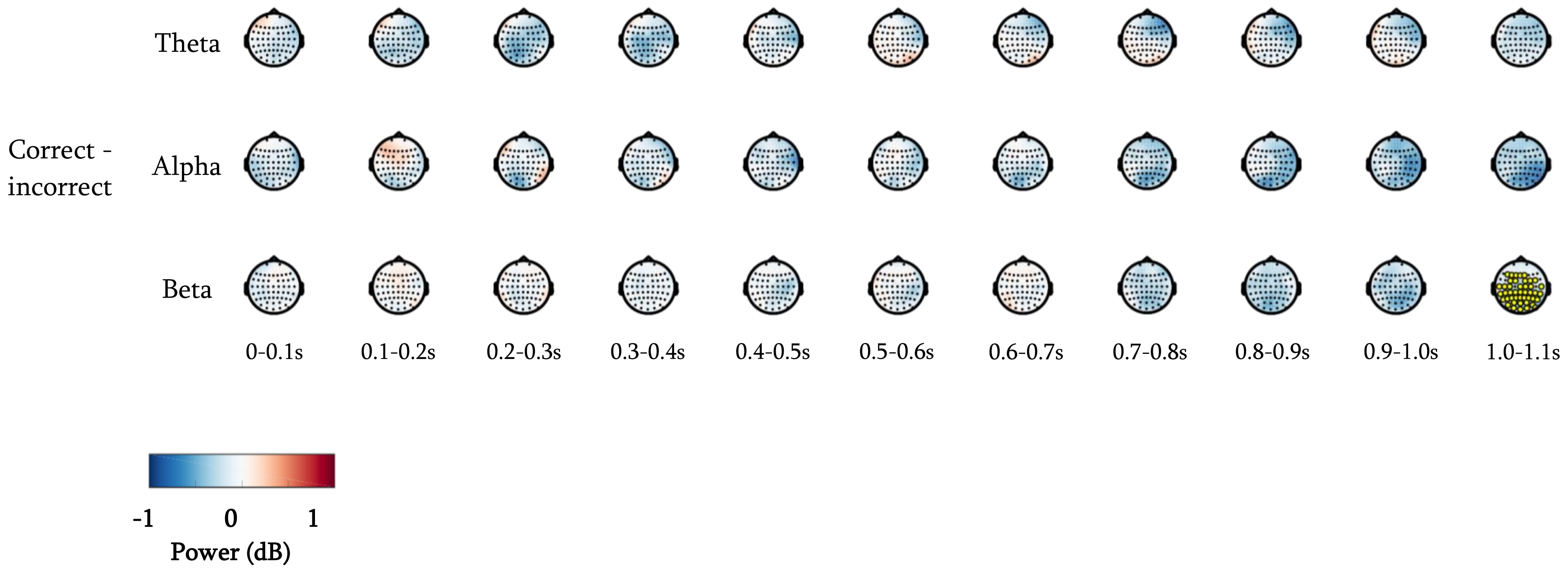

Figure 2.4. Example cluster plot of time-frequency data. Topographical plots (showing power differences between two conditions), with time and electrode data samples that were part of a significant cluster highlighted by a yellow dot.

In summary, the present methodology chapter provides an introduction to the key methods used throughout the current thesis. Next, four experimental chapters will follow that report on experiments that investigated retrieval-induced updating of face memories. Within these chapters, methodological information regarding the topics discussed in the present chapter will include more specific details (e.g., parameters used in creating face stimuli, details of statistical tests, etc.) that are relevant to specific chapter content.

Chapter 3 - Retrieval-induced updating for face memories is modulated by objective and subjective retrieval processes

The memorial consequences of repeated retrieval have previously examined the role of both objective and subjective retrieval processes leading to memory updating (Marsh et al., 2007; Roediger & Marsh, 2005; St. Jacques et al., 2013). As mentioned in Chapter 1, research has suggested that enhanced subjective retrieval processes during retrieval can both be associated with improved later memory for information correctly retrieved, but also an increased likelihood of updating memories with false information (St. Jacques et al., 2013; St. Jacques & Schacter, 2013). A relevant interpretation for these effects suggest that neural reactivation of memories during retrieval, a critical factor to induce memory updating (Hupbach et al., 2007; 2008; 2009; Sederberg et al., 2011), was more likely for memories retrieved with enhanced subjective experience of remembering. Whereas previous research has examined how objective and subjective retrieval processes interact to modulate retrieval-induced updating of more complex 'relational' episodic memories involving associations between multiple different items and features (such as object-location associations or autobiographical memories), the updating mechanisms of simpler item memories is unknown.

Recognition of items has been shown to involve qualitatively different neural mechanisms to recollection of relational memories (Ranganath & Ritchey, 2012). Item recognition might be achieved through a familiarity processes that is associated with less neural reinstatement compared to recollection (Johnson, McDuff, Rugg & Norman, 2009). Therefore, updating of item memories via recognition may be achieved through different mechanisms to those established in previous work (Bridge & Paller, 2012; St. Jacques et al., 2013). The experiments in the present chapter aimed to find behavioural evidence of how objective recognition success and subjective recognition awareness predicts item memory updating, using face stimuli as item memories.

In these studies, recognition confidence was used as a measure of subjective retrieval experience. With regards to episodic memory retrieval, a simplistic view of how recognition accuracy and confidence correspond during retrieval suggests that memories that are accurately recognised should be associated with elevated confidence. However, the relationship between confidence and accuracy has been shown to depend on the design of episodic memory tasks. For example, a collection of studies by Roediger and DeSoto (DeSoto & Roediger, 2014; Roediger & DeSoto, 2014) demonstrated that negative/null correlations between accuracy and confidence emerged when participants were judging whether they recognised new words that were semantically related to old items, driven by higher confidence during false recognition of the semantically related lure items. This result can be accommodated by models of memory that considers confidence responses during recognition as a correlate of the strength of the recognition signal that the memory evokes; an assumption from signal detection theories of recognition memory (Busey et al., 2000; Parks & Yonelinas, 2007; Thakral et al., 2015) and face space models (Lewis, 2004). That is, a stronger recognition signal corresponds to higher confidence judgements, not only for true memories but also for false recognition of new items that are similar to stored episodic memories (e.g., related lures in DeSoto & Roediger, 2014; Roediger & DeSoto, 2014). It could be argued, therefore, that elevated confidence during recognition is indicative of stronger memory reactivation and is therefore a marker of those item memories that are more likely to be updated during retrieval.

Two experiments were conducted in Chapter 3. The first experiment (3a) aimed to examine whether elevated confidence responses during initial recognition attempts led to biased recognition responses during Test 2. The second experiment (3b) had the same aims but modified the confidence response scale (see section 3.2). Both experiments used face stimuli artificially generated from face space models (Solomon et al., 2015; see Method section 2.1). The recognition task used in both experiments presented target faces (shown during a previous learning task) amongst several distractor faces. This paradigm allowed the examination of whether correct and incorrect responses during initial recognition would be repeated during Test 2, as a function of confidence in responses during Test 1.

3.1. Experiment 3a

In a novel repeated recognition paradigm, participants first completed a learning phase where 30 face targets were encoded in three cycles (see Appendix A for validation experiment). Following learning, participants completed a recognition task for 20 of the 30 face targets in Test 1. In each trial, a target face was shown with four distractor faces from the same face set. Participants were asked which face they thought was the target learnt previously. Subjective measures for each trial were also acquired; whether participants reported high or low confidence in their recognition decision, and for low confidence responses, whether this was due to experiencing recognition of ‘multiple’ faces, or ‘no’ recognition of any of the faces. The same recognition task was completed at Test 2 recognition for the 20 face sets shown during Test 1 recognition (repeated trials) and the 10 face sets not shown during Test 1 (baseline trials). From this design, accuracy for trials completed during Test 1, as well as accuracy for repeated and baseline trials completed during Test 2 were acquired. At Test 2, face recognition ‘bias’ was measured for repeated trials; indicating whether participants repeated the same recognition choices made during Test 1.

Relevant to the central hypothesis of this experiment, it was expected that recognition bias during Test 2 would be modulated by the accuracy and confidence of recognition responses during Test 1 (St. Jacques et al., 2013; St. Jacques & Schacter, 2013). That is, faces recognised during Test 1 with high confidence were expected to be re-selected at Test 2, compared to Test 1 responses made with low confidence; consistent with the argument that retrieval promotes the encoding of face images selected at Test 1 that then biases recognition responses in Test 2 (Bridge & Paller, 2012). The same effect of confidence on bias was expected for both correct and incorrect Test 1 responses, consistent with the view that retrieval can be both beneficial and detrimental to subsequent remembering (Butler et al., 2006; Marsh et al., 2007; Roediger & Marsh, 2005).

3.1.1. Method

Participants

Thirty-eight participants aged 18-21 (M_{age} = 19.21, SD_{age} = 0.93), with 37 females, completed the experiment in a within-subjects design. This sample size was recruited to maximise participant numbers during a limited recruitment time window. However with this sample size, there was 85% power to detect a medium effect size (0.5) and 44% power to detect a small effect size (0.3) with a two-tailed alpha of .05. Participants were all Psychology students at the University of Kent and completed the experiment in return for course credit. Exclusion criteria prevented participants from taking part if they were not aged 18-35, or if they had taken part in any other experiment from this thesis. All participants had normal/corrected to normal vision. Ethical approval was acquired from University of Kent, School of Psychology ethics committee.

Stimuli

Face stimuli. Face stimuli used for the present experiment have been previously described in Chapter 2. As a reminder, face stimuli were sampled from a face space constructed of 29 components. From this model, 60 face sets were created, each face set containing five faces (images 1-5). Initially, 60 face locations were sampled at a constant radius from the face space origin, with target to target Euclidean distances ranging from ~4000 - ~7000. For these 60 initial face locations, four additional face locations were randomly sampled from a uniform distribution around the initial face, with a limit of +/- four standard deviations. The perceptual dissimilarity between the five face images within each set corresponded to a larger Euclidean distance (see Appendix B for validation experiment). From each set of five face images, one face image was designated as the ‘target face’ shown during learning and recognition tests, whilst the other four faces were ‘distractor faces’ shown only during recognition tests. Assignment of face images to the target condition was counterbalanced across participants, so that each of the face images (images 1, 2, 3, 4 and 5) were set as the target face an equal number of times across participants.

Apparatus

The experiment was completed on a Dell optiplex 9020 desktop computer with PsychoPy (Peirce, 2009). The screen measured 51.0cm x 28.4cm, with a resolution of 1920 x 1080 pixels. All participants completed the experiment on the same computer in the same laboratory.

Design and Procedure

Participants completed two blocks, with 30 face sets per block and each block containing the same learning, filler task, recognition Test 1, filler task and recognition Test 2 format. Block one presented face sets 1-30 and block two presented face sets 31-60 (see Figure 3.1 for schematic of the procedure).

Learning. Participants completed three learning cycles, with all 30 face targets shown in each of the three cycles. For each cycle, face targets were presented one at a time in a randomised order for 5000ms each, preceded by a 500ms fixation. All face pictures subtended a 5.12 x 5.88 visual angle when participants sat at a distance of 75cm from the screen. During face presentation, participants were instructed to memorise the uniqueness of each face image for the later memory tests. Additionally, participants completed a rating task for each face target in order to facilitate encoding. During cycle one, participants decided whether the person of each face was nice (key press 5) or not nice (key press 1). Cycle two involved rating faces as attractive (key press 5) or unattractive (key press 1). The third cycle required participants to indicate whether a face was female (key press 5) or male (key press 1). Participants were informed of each rating task prior to beginning each cycle and made their responses whilst the face was on screen.

Letter search filler task. Following learning, participants completed the Treisman and Gelade's (1980) visual letter search task as a filler task. The purpose of this task was to separate memory encoding and recognition tasks so that performance during the recognition task was based on long-term rather than short-term working memory. Seventy-two letter arrays which contained a mixture of letter type (X, O & N), frequency (1, 5, 15 & 30) and colour (red, green & blue) combinations in each picture were presented one at a time. Participants were tasked with searching for a blue letter, which was present in 18/72 pictures. Participants indicated whether the array had a blue letter (key press 5) or did not have a blue letter (key press 1). Trials were self-paced, with all 72 trials completed in around one minute.

Test 1 recognition. The Test 1 recognition task involved 20/30 of the face sets, with one set shown per trial. Face sets that were assigned to the repeated condition were counter-balanced across participants. For each trial, one target face was presented with the four distractor faces from the corresponding face set. These five faces were presented sequentially, numbered from 1-5, and shown for 1250ms each with a 500ms fixation preceding each face. Faces were presented sequentially to ensure that participants focused on encoding all faces within a trial (rather than just focus on a handful of faces if presented simultaneously, which would be problematic as there would be much more variability in the extent to which participants focused on targets/distractor faces across trials). Targets were equally likely to appear in each one of the five positions in the sequence (i.e. 1/5 of targets were shown first, 1/5 second, etc.). All face images subtended a visual angle of 5.12 x 5.88, and each face image was located in each of the five positions an equal number of times. After viewing all five faces, participants were shown with the numbers 1-5 in the centre of the screen, and were asked to respond and indicate which face they thought was the target face shown during learning, using a keyboard press (1, 2, 3, 4 or 5). Participants had a maximum of 6000ms to respond, with a question mark prompting a response at 3000ms. After making their recognition decision, participants were asked to rate their confidence that the face they picked was the target (1 = *Definitely unsure*, 2 = *Slightly unsure*, 4 = *Slightly sure*, 5 = *Definitely sure*), with responses self-paced. If participants reported either ‘*Definitely unsure’* or *‘Slightly unsure’* confidence, a 'reason' question was presented to find out why participants reported having low confidence (though reason responses were not required for high confidence judgements as it was decided that participants would not have multiple possible reasons for providing a high confidence judgement). Here, participants either reported having ‘no memory’ for the target face (key press 1) or reported recognising ‘multiple faces’ as the target (key press 5), as it was assumed that low confidence could stem from a failure to recognise any face, or from recognising several faces and not being able to tell which face was the target. Reason responses were self-paced. Termination of the reason question preceded a 1000ms intertrial interval, following which the next face set was presented. The trial structure continued until all 20 trials were completed. Face set order was randomised for each participant, therefore the order of the location of target faces/correct responses was random across trials.

Letter search filler task. The same letter search task separated the end of the Test 1 recognition and start of Test 2 recognition. Letter array order was randomised and again took around one minute to complete.

Test 2 recognition. Test 2 recognition involved the same trial format as described in the Test 1 recognition task, with 30 trials completed. Of these, 20 were ‘repeated trials’ containing the same sets of five images that were shown during Test 1, and 10 ‘baseline’ trials that were not shown during Test 1. For face sets in the repeated condition, the order of the five faces within each trial was randomly reshuffled from the Test 1 trial. The 30 recognition trials were presented in a random order, after which participants had a short break before completing the second block.

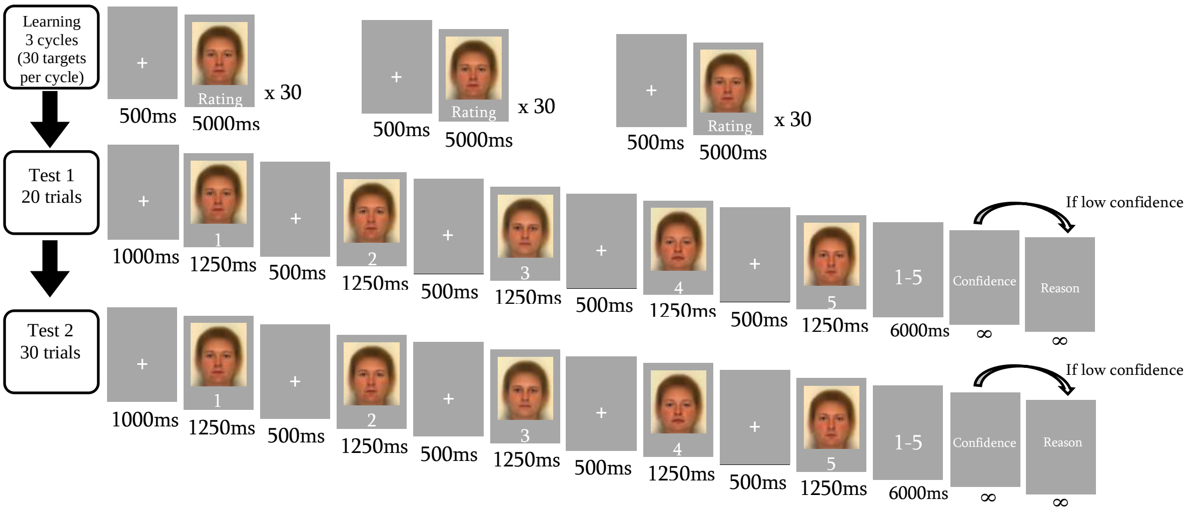

Figure 3.1. Procedure for Experiment 3a. During learning, participants encoded 30 face targets. After a short filler task, participants completed two recognition tests where they were asked to identify target faces amongst four distractor faces. Confidence and reason for low confidence responses were also collected.

Data analysis

Data analysis was performed on data collapsed across block one and two to maximise statistical power. Initially, omnibus measures of recognition accuracy, confidence and reasoning for low confidence responses were calculated for Test 1 and Test 2 recognition trials. Test 1 accuracy was calculated as the proportion of Test 1 trials where participants correctly selected the target face. Test 2 accuracy for repeated trials was calculated as the proportion of repeated trials where the target face was selected at Test 2. Finally, Test 2 accuracy for baseline trials was calculated as the proportion of baseline trials where the target face was recognised at Test 2 recognition. Measures of subjective recognition were also analysed for each condition. For Test 1, Test 2 repeated and Test 2 baseline conditions, proportion of high confidence responses was calculated. Further, for low confidence trials in Test 1, Test 2 repeated and Test 2 baseline, the proportion of ‘multiple memory’ responses (as opposed to ‘no memory') was also calculated.

In addition, a Test 2 recognition bias measure was calculated to determine the extent to which participants were biased by prior Test 1 recognition choices during Test 2. The Test 2 recognition bias measure was calculated separately according to accuracy and confidence of Test 1 recognition responses. Within participants, Test 2 recognition trials in the repeated condition were categorised according to whether responses in the previous Test 1 trial were correct or incorrect. Both correct and incorrect trials were further split according to the confidence of the Test 1 response; high confidence was categorised as a response of *slightly sure* or *definitely sure*, whereas low confidence was categorised as a response of *slightly unsure* or *definitely unsure*, leaving conditions of Correct Test 1 High Confidence, Correct Test 1 Low Confidence, Incorrect Test 1 High Confidence and Incorrect Test 1 Low Confidence [1].

For “prior error” trials, the proportion of these trials where a different distractor was selected during Test 2 was also calculated, with these non-bias values were divided by three to calculate the proportion of trials that participants selected one of the three possible non-bias distractors during Test 2 recognition. By deducting the non-bias proportion values from the proportion bias measures, this “bias difference score” estimated how much more likely participants were to repeat prior Test 1 recognition responses during Test 2 following high and low confidence recognition decisions from Test 1. The bias difference score for prior errors were analysed to show how the original proportion bias measures and the bias difference scores produce similar outcomes, validating the bias difference score for the analysis of memory updating for the experiments in Chapter 4 (see Appendix C.3

---

[1] Only two categories of high vs low confidence were created as there were insufficient trial numbers to analyse each of the individual four response categories.

for these results).

3.1.2. Results

Influence of repeated retrieval on objective and subjective recognition

The first analyses focused on how repeated retrieval effects general patterns of face recognition accuracy, confidence and reasons for low confidence responses. For all Test 1, Test 2 repeated and Test 2 baseline trials, the proportion accuracy, proportion high confidence and proportion of multiple memories (for low confidence trials) were calculated (see Figure 3.2 for descriptive statistics). Within each measure, a one-way ANOVA was conducted to compare initial, repeated and delayed recognition on accuracy, confidence and multiple memory responses.

Proportion accuracy. For the proportion accuracy measure, the one-way ANOVA was not significant (*F*(1.68,62.00) = 1.69, *p* = .20, η^2p = 0.04). Thus, whereas mean recognition accuracy seems to be higher for Test 1 compared to Test 2 conditions, no significant effects between conditions emerged.

Proportion confidence. For the confidence measure, a one-way ANOVA was also not significant (*F*(1.57,57.92) = 3.24, *p* = .06, η^2p = 0.08) showing confidence was not significantly different between the initial Test 1 responses compared to responses when those sets were repeated at Test 2[2], or sets in the baseline condition.

Proportion multiple memories. For proportion of low confidence responses that were associated with an experience of recognizing multiple faces, the one-way ANOVA was significant (*F*(1.67,61.67) = 4.68, *p* = .02, η^2p = 0.11). Following this, paired samples

---

[2] Follow-up analysis to explore why confidence was descriptively higher in Test 2 for repeated trials showed that trials where participants were correctly recognising target faces across both tests had more high confidence judgements compared to all other accuracy/error combinations between Test 1 and 2, see Appendix C.1.

t-tests (Bonferroni corrected α = .016) indicated that reports of multiple memory responses were not different between Test 1 trials and Test 2 repeated conditions (*t*(37) = 1.38, *p* = .18, *g* = 0.22). Furthermore, the proportion of multiple memory responses was not greater for Test 1 compared to Test 2 baseline (*t*(37) = 1.75, *p* = .09, *g* = 0.28). In contrast, participants reported significantly more multiple memory experiences in the repeated condition compared to the baseline condition during Test 2 (*t*(37) = 2.80, *p* = .008, *g* = 0.45). These results thus show that participants experienced low confidence due to recognising more than one face more often during Test 1 and for repeated trials during Test 2 compared to the baseline condition in Test 2.

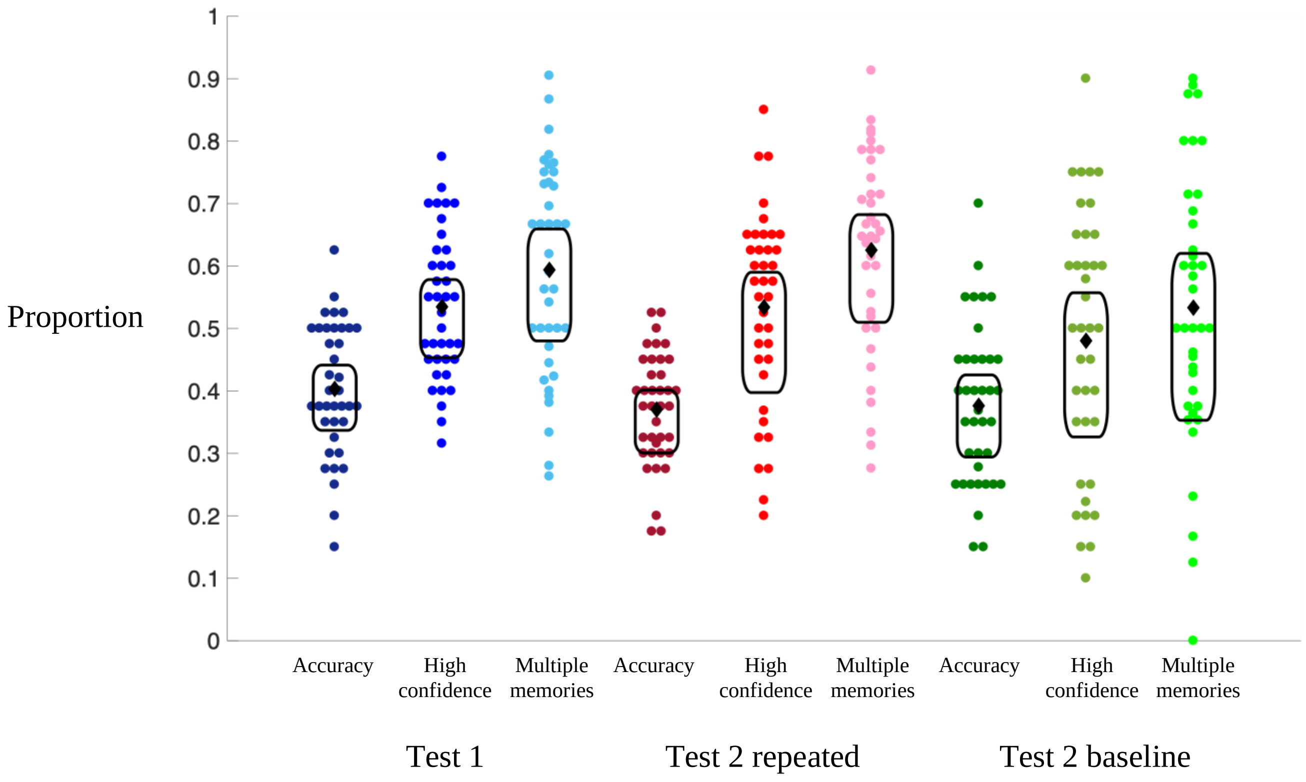

Figure 3.2. Behavioural data of accuracy, confidence and reason for Experiment 3a. Proportion accuracy, high confidence and 'multiple memory' responses for Test 1, Test 2 repeated and Test 2 baseline conditions.

Accuracy and confidence of initial recognition modulates face memory updating

The key analysis for Experiment 3a examined patterns of recognition bias during Test 2 as a function of accuracy and confidence of recognition during Test 1. As mentioned, Test 2 recognition bias was quantified according to whether participants were more likely to repeat the same recognition choice from Test 1 to Test 2, a.k.a proportion bias. This measure was statistically analysed when the preceding Test 1 trial was correct made with high confidence versus correct made with low confidence (a.k.a correct bias). Similarly, proportion bias was statistically analysed when the preceding Test 1 trial was incorrect made with high confidence versus incorrect made with low confidence (a.k.a incorrect bias). It was reasoned that proportion bias should be larger for high confidence versus low confidence conditions, for both correct and incorrect Test 1 trials, supporting the argument that high confidence recognition during Test 1 promotes the encoding of recognition responses that bias subsequent recognition.

Exclusion criteria for the following analyses were applied to ensure that participants had sufficient trial numbers for each of the Test 1 correct high, correct low, incorrect high and incorrect low conditions, in order to calculate reliable proportion bias measures. Participants with fewer than 5 trials in either correct high or low conditions were removed, leaving an analysis sample of 23 participants for the correct bias analysis. Additionally, participants with fewer than 5 trials in either incorrect high or low conditions were removed for the incorrect bias analysis, leaving an analysis sample of 37 participants.[37]To check that removing participants did not bias the results, the following analyses were conducted with the inclusion of participants who did not have enough trial numbers. The outcome of these analyses did not change the interpretation of the results, meaning that removing participants with insufficient trial numbers was not likely to bias these analyses towards what was hypothesised. Therefore, the following analyses were conducted with the exclusion of participants who did not provide enough trials for each condition. Note that, with these smaller sample sizes, there was there was a reduced statistical power of 63% and 28% power to detect medium (0.5) and small effect sizes (0.3) for the correct bias analysis, and 84% and 43% power to detect medium (0.5) and small effect sizes (0.3) for the incorrect bias analysis.

Correct bias. Descriptive statistics for the correct bias analysis can be seen in Figure 3.3. Proportion bias measures during Test 2 were significantly greater than chance

---

[3] In an alternative analysis (not presented), the same analyses were repeated after also removing participants who had less than 0.2 accuracy during refresh retrieval (i.e. those who show poor memory for targets). Two participants were identified and removed for the correct bias analysis, however the conclusions from this analysis are the same as those in-text, therefore participants were included even if they had low test 1 accuracy to avoid introducing biases in the sample.

performance of 0.2 (correct high - *t*(22) = 8.12, *p* <.001; correct low – *t*(22) = 4.10, *p* <.001). However, a paired samples t-test showed that proportion bias was significantly greater for the correct high compared to the correct low condition (*t*(22) = 2.09, *p* = .05, *g* = 0.44), suggesting that participants were more biased during Test 2 when previous correct responses were made with high confidence.

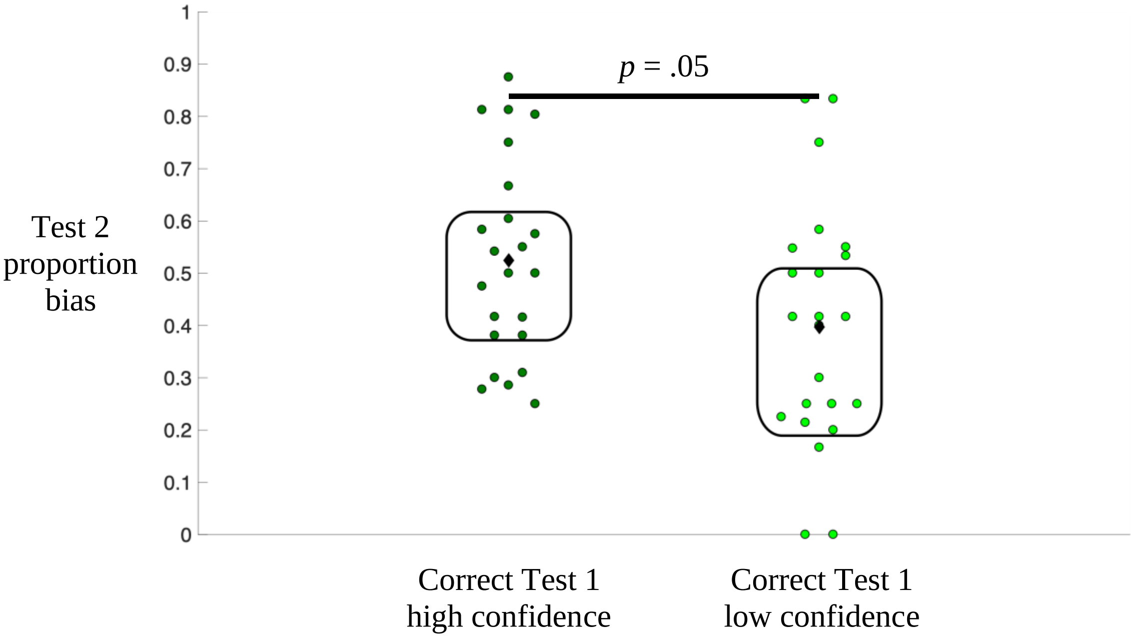

Figure 3.3. Proportion bias following Test 1 correct high and low confidence. Test 2 proportion bias for Test 1 correct high and Test 1 correct low. Participants were more likely to repeat target selections during Test 2 recognition when previous target recognition was made with high versus low confidence.

Incorrect bias. Next, the proportion bias measure was analysed for the Test 1 incorrect high and low confidence conditions (see Figure 3.4 for descriptive statistics). The proportion measure was significantly greater than chance performance of 0.2 for the incorrect high condition (*t*(36) = 4.70, *p* <.001), however, proportion bias was no higher than chance performance for the incorrect low condition (*t*(36) = 0.53, *p* = .60). Furthermore, proportion bias was significantly higher following incorrect high versus incorrect low conditions (*t*(36) = 4.21, *p* <.001, *g* = 0.69). These results are consistent with the prediction that participants were more likely to be biased by previous recognition errors during Test 2 when prior errors were made with high compared to low confidence.

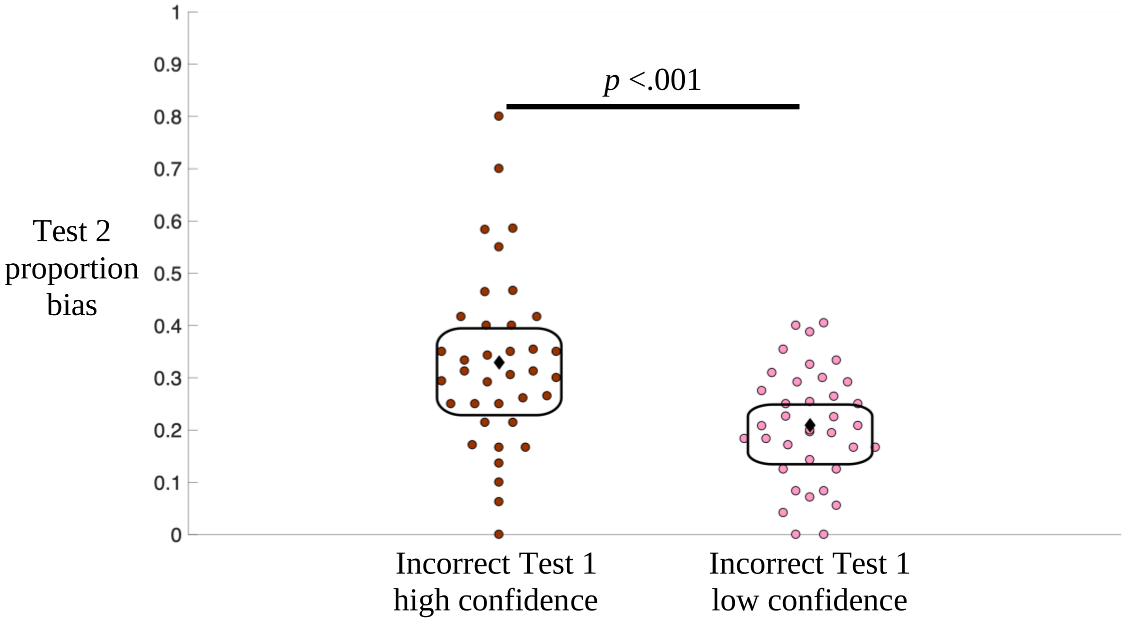

Figure 3.4. Proportion bias following Test 1 incorrect high and low confidence. Test 2 proportion bias following incorrect high and incorrect low trials. Participants were more likely to repeated incorrect retrieval during Test 2 when previous recognition errors were made with high versus low confidence.

Relationship between Test 1 accuracy and Test 2 bias. The outcome of the previous analysis demonstrated that Test 2 recognition performance was influenced by the subjective retrieval processes from prior Test 1 recognition attempts. It was next queried whether certain types of recognition responses during Test 1 were more likely to be repeated during Test. Similar analyses were done by Bridge and Paller (2012), who found that Test 2 retrieval bias did not depend on the distance between object-locations remembered during Test 1 and the original object location (i.e. no relationship between Test 1 accuracy and Test 2 bias). Similarly, this analysis capitalised on the strength of using the continuous Euclidean distance as a measure of the extent to which participants recognised faces during Test 1 that were more similar to the target, and whether this recognition precision influenced subsequent memory bias in Test 2. For each participant, the Euclidean distance between the face selected at Test 1 and the target face was acquired for each trial (Test 1 error). In addition, for each trial the Euclidean distance between the face selected at Test 1 and the face selected at Test 2 was acquired (Test 2 bias). Pearson's correlations between Test 1 error and Test 2 bias were calculated, collapsed across block (i.e., the correlation was performed on all 40 repeated trials). Across participants, the average correlation coefficient between Test 1 error and Test 2 bias (r = .29) was significantly greater than zero (t(37) = 12.03, p <.001, g = 1.95). As seen in Figure 3.5, participants consistently showed a positive relationship, though varying in magnitude, between the size of recognition errors made at Test 1 and selecting a similar face at Test 2. The same pattern of results appears when only taking into account Test 1 trials where a distractor face was selected (r = .32; t(37) = 11.30, p <.001, g = 1.83; Figure 3.5). These results imply that, for the majority of participants, recognition responses during Test 2 were biased more towards Test 1 recognition decisions when the face recognised during Test 1 was the same/similar to the target face.

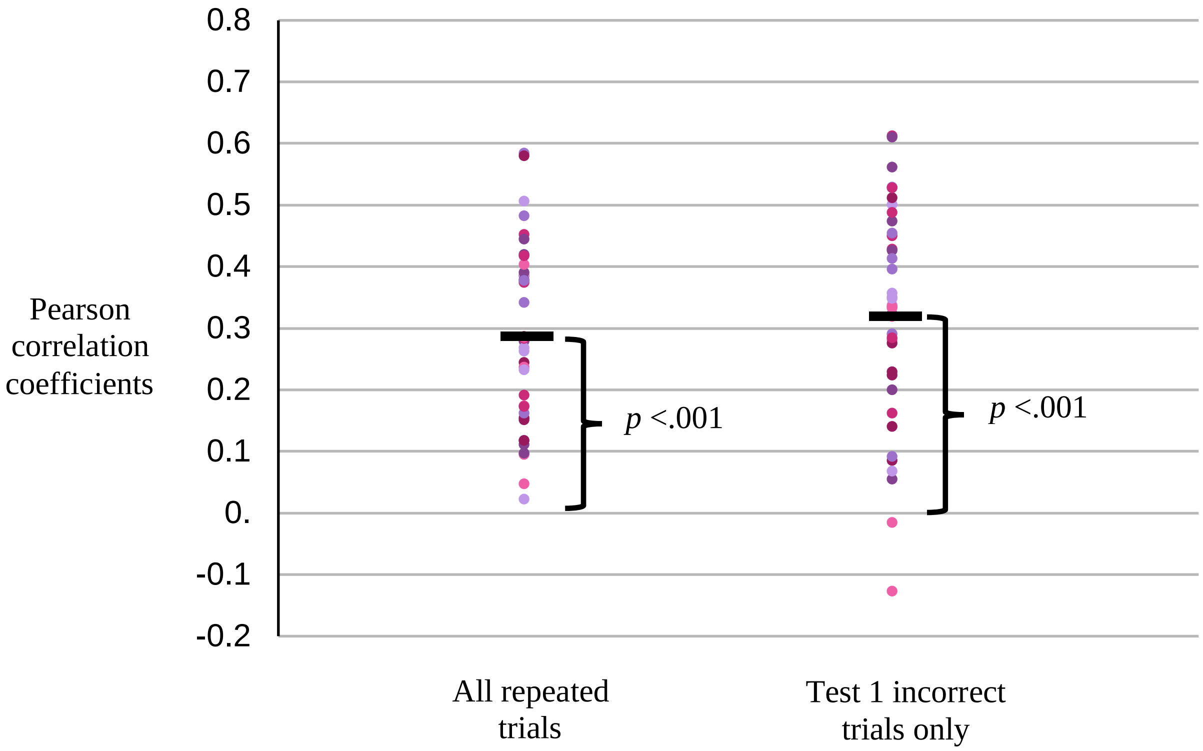

Figure 3.5. Correlation between Test 1 accuracy and Test 2 bias. Pearson correlation coefficients between Test 1 recognition error and Test 2 bias for all repeated trials and incorrect Test 1 trials. Mean correlation coefficients (horizontal black bars) suggest that Test 2 recognition was similar to Test 1 recognition responses when initial recognition was more similar to targets.

3.1.3. Discussion

The aims of the first experiment of Chapter 3 were two-fold; 1) to compare objective and subjective retrieval measures across repeated and delayed recognition, 2) to examine whether confidence during initial recognition related to face memory updating, as indicated by biased responses during subsequent retrieval attempts. Relevant to the second aim, the present experiment showed that recognition responses made with high confidence during Test 1 were more likely to bias subsequent retrieval attempts, irrespective of whether the initial recognition judgement was correct or incorrect. These results are consistent with previous findings showing both beneficial and harmful effects of repeated retrieval, and suggest that such effects may be modulated by recognition confidence during initial retrieval (St. Jacques et al., 2013; St. Jacques & Schacter, 2013).

Furthermore, and relevant to the first aim, data from Experiment 3a showed objective and subjective retrieval processes were differentially affected by repeated and delayed recognition (see Figure 3.2). For retrieval accuracy, participants appeared to have better memory for targets during Test 1 compared to Test 2 repeated and baseline conditions, though these differences were not statistically reliable. Subjective retrieval across both tests were measured according to recognition confidence, as well as providing a reason (recognising multiple faces per trial vs. recognising none of the faces per trial) for low confidence responses. Recognition confidence appeared to be higher for Test 1 and Test 2 repeated conditions compared to the Test 2 baseline condition, however again these differences were not statistically reliable. Finally, proportion of 'multiple memory' responses were shown to be larger during Test 1 and Test 2 repeated conditions compared to Test 2 baseline. Collectively, these findings suggest how repeated retrieval differentially influenced objective and subjective recognition processing of face memories.

Although the results of Experiment 3a shows some interesting initial results regarding the influence of repeated recognition testing on face recognition accuracy, confidence and updating, one limitation with this study concerns the confidence scale used. Participants indicated their recognition confidence on a four-point Likert scale, which led to many participants producing only a small number of low confidence correct responses during Test 1, thus reducing the possibility for participants to repeat low confidence correct responses between tests. This aspect of the design led to a large number of participants being excluded from key analyses. Consequently, a follow-up study was conducted with an amended ‘continuous’ confidence scale. This confidence scale allowed high versus low confidence conditions to be computed within participants relative to their median confidence score, instead of enforcing categorical options across all participants such as the scale used in Experiment 3a. Therefore, nearly all participants could contribute a sufficient number of trials to all conditions for reliable measurement of performance.

3.2. Experiment 3b

Experiment 3b was near identical to Experiment 3a, with participants completing two blocks of learning, filler, Test 1 recognition, filler and Test 2 recognition tasks. In both blocks, participants learnt 30 target faces before completing a recognition task for 20 of the 30 face targets. For each trial, a target face was shown with the distractor images from the same set as the target. Participants indicated which of these faces they recognised from the learning phase, and provided a confidence response for their recognition decisions. The same recognition task was completed in Test 2 but for the 20 face sets shown during Test 1 (repeated trials) and the 10 face sets not shown during Test 1 (baseline trials). The main difference from Experiment 3a was the confidence scale. In Experiment 3b, participants made keyboard responses with varying keypress lengths to indicate higher and lower confidence (longer key press associated with higher confidence). High and low confidence conditions were created from this continuous measure by performing a median split of keypress length. Consequently, an even number of high confidence and low confidence responses could be used for analysing patterns of retrieval-induced updating.

As these confidence responses were relative and not clearly linked with verbal labels, it was no longer possible to ask participants for their reason for low confidence responses (i.e. whether they experienced multiple vs. no memory), therefore this question was removed. As a reminder of the hypotheses, and consistent with results from

Experiment 3a, it was hypothesised that high confidence responses during Test 1 would be more likely to be repeated at Test 2, both when Test 1 recognition was correct and incorrect, reflecting the influence of subjective retrieval processes on face memory updating. Furthermore, it was expected that face recognition accuracy would be greater during Test 1 compared to Test 2, similar to those results found in Experiment 3a.

3.2.1. Method

Participants

Fifty-four student participants (46 female) with an age range of 18-29 were recruited from the University of Kent (M_{age} = 19.91, SD_{age} = 3.53). This sample size was recruited to maximise participant numbers during a limited recruitment time window. However with this sample size, there was 95% power to detect a medium effect size (0.5) and 58% power to detect a small effect size (0.3) with a two-tailed alpha of .05. Participants received course credit in return for participation, in line with ethical approval from the University of Kent School of Psychology ethics board. Exclusion criteria prevented participants from taking part if they were not aged between 18-35, did not have normal/corrected-to-normal vision (including if participants were red-green colour blind) or were recruited for any other experiment from the current thesis, meaning all participants were naïve to the aims of the study.

Stimuli

Face stimuli. The same face stimuli from Experiment 3a were used in Experiment 3b (see Experiment 3a methods).

Apparatus, procedure and design

The apparatus and procedure for Experiment 3b were identical to Experiment 3a, except for the confidence scale used during recognition tasks in Test 1 and 2, and the removal of the “reason” question for low confidence responses (see Figure 3.6). Participants completed two blocks, both blocks starting with a learning phase where 30 targets were encoded. Following a short letter search filler task, participants completed 20 recognition trials in Test 1. For each trial, a target face was shown amongst four distractor faces, with participants required to select the face they recognised from the previous learning phase, and provide a judgement of how sure they were that the target face was chosen. Following a second letter search filler task, participants completed the same recognition task in Test 2, for 20 repeated trials (shown during Test 1) and 10 baseline trials (not shown during Test 1).

For Experiment 3b, the confidence scale used presented the numbers 1-5 in white font (RGB colour Red = 1, Green = 1, Blue = 1) on screen. Participants were instructed to press and hold down the key associated with the face they thought was the target face, amongst the five faces just seen which combined both the recognition and confidence judgements into a single, time efficient method for trial by trials responses. Once a key was pressed, the number on-screen corresponding the pressed key changed to green (Red = -1, Green = 1, Blue = -1). The colour of the number then gradually changed from green to red whilst the pressed key was held down, with the maximum colour change reached after ~4 seconds (Red = 1, Green = -1, Blue = -1). That is, the longer a key was held down for, the more red the number became on screen and it was explained to participants that red indicated higher confidence and green indicated lower confidence. Once the key was released, the next trial began. Participants were instructed to use the whole range of the colour scale to indicate different degrees of confidence, and were given two practice trials prior to the start of the test trials.

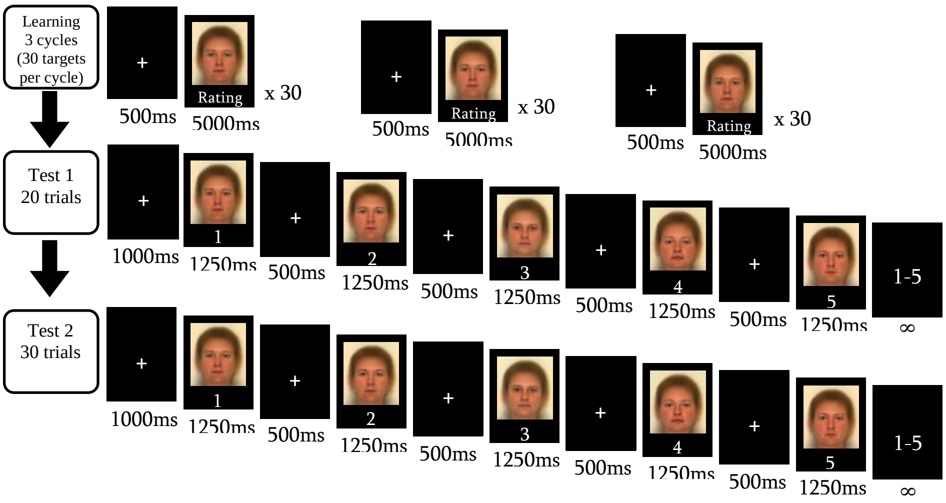

Figure 3.6. Procedure for Experiment 3b. During learning, participants encoded 30 face targets. After a short filler task, participants completed two recognition tests where target faces were identified amongst four distractor faces, with confidence responses provided with a continuous confidence scale.

Data analysis

As in Experiment 3a, the same measures of recognition accuracy and bias were calculated, separately for blocks one and two. Three exceptions were made to Experiment 3b in consideration of the change on confidence scale used. To analyse changes in recognition confidence between conditions in Experiment 3b, the mean keypress durations (in seconds) from trials within each condition was calculated. In addition, within Test 1 trials, Test 2 repeated and Test 2 baseline conditions, the relationship between accuracy and confidence was quantified by calculating Pearson's correlations between the size of recognition errors and keypress durations during each recognition trial. Finally, to analyse Test 2 recognition bias as a function of Test 1 accuracy and confidence, relative high and low confidence conditions were calculated within-subject, separately for Test 1 correct and Test 1 incorrect trials. When the Test 1 trial was correct, the median of keypress durations during the Test 1 trial was calculated, with trials with keypress durations above the median allocated to the correct high confidence condition, and trials with keypress durations below the median allocated to the correct low confidence condition. Similarly, the median of keypress durations made during Test 1 was calculated for incorrect Test 1 trials, with trials with keypress durations above the median allocated to the incorrect high confidence condition, and trials with keypress durations below the median allocated to the incorrect low confidence condition. The median split procedure meant that correct and incorrect conditions were not directly comparable (as different medians were calculated for correct and incorrect trials separately). However, this analysis was interested in comparing bias for high vs. low confidence conditions within correct and within incorrect trials, rather than to compare bias measures between correct and incorrect trials (as it was expected that bias would be larger for correct vs. incorrect trials anyway, based on data from Experiment 3a as well as the fact that target faces presented during learning and recognition would evoke much stronger recognition than distractor faces presented only during recognition). However, to confirm that participants' confidence responses were being adequately allocated to high and low confidence conditions respectively, the mean confidence of responses was shown to be higher for the correct high versus correct low confidence, as well as the incorrect high versus incorrect low confidence conditions (see Appendix D.3).

One novel aspect of data analysis for Experiment 3b was to incorporate both Bayesian and frequentist statistical analysis. One strength of Bayesian analysis is that this approach shows the ratio of evidence in favour of one hypothesis over another from observed data. For each t-test performed in the following analysis, Bayes factors were calculated with informed priors, these being the observed effect sizes seen for the equivalent contrast shown in Experiment 3a. By adding Bayes Factor statistics, the data from Experiment 3b was able to not only indicate whether effects were significantly different, but also indicated how much the data favoured the alternative versus null hypotheses.

3.2.2. Results

Influence of repeated retrieval on objective and subjective recognition

The initial analysis of Experiment 3b compared whether recognition accuracy and confidence varied between Test 1, Test 2 repeated and Test 2 baseline trials. Proportion accuracy (was the target face selected) and mean confidence (mean keypress durations from confidence responses) measures were analysed to assess the influence of repeated and delayed retrieval on objective and subjective recognition processes.

Proportion accuracy. The proportion accuracy measures for Test 1 ($M = 0.37$, $SE = 0.01$), Test 2 repeated ($M = 0.33$, $SE = 0.01$) and Test 2 baseline conditions ($M = 0.32$, $SE = 0.02$) were entered into a one-way repeated measures ANOVA, which was significant ($F(2,106) = 7.19$, $p = .001$, $\eta^2 p = .12$). Statistical comparisons within this effect (corrected $\alpha = .016$; see Table 3.1 for inferential statistics) established that Test 1 accuracy was significantly greater than Test 2 recognition accuracy for both repeated and baseline conditions. However, Test 2 recognition accuracy was not significantly different between repeated and baseline conditions. Furthermore, Table 3.2 shows how Bayes factors confirm evidence in favour of the alternative versus null hypothesis for both the Test 1-Test 2 repeated and Test 1-Test 2 baseline contrast, supporting the hypothesis that recognition accuracy would be larger during Test 1 vs. Test 2. Furthermore, the Bayes Factor for the Test 2 repeated-baseline contrast showed more evidence in favour of the null vs alternative hypothesis. In summary, these results are consistent with the descriptive data from

Experiment 3a, however with an increased statistical power these results show the statistical reliability of accuracy being largest during Test 1 compared to Test 2, with no difference between repeated and baseline conditions during Test 2.

Table 3.1. Inferential statistics for proportion accuracy analysis in Experiment 3b.

Condition contrast	df	*t*	*p*	*g*	BF	Prior
Test 1 – Test 2 repeated	53	3.48	<.001	0.47	57.81	0.31
Test 1 – Test 2 baseline	53	3.30	<.001	0.45	34.78	0.19
Test 2 repeated – Test 2 baseline	53	0.66	.51	0.09	0.19	0.06

Mean confidence. For confidence measures, the descriptive statistics suggest the confidence for Test 1 (*M* = 1.79, *SE* = .10) was larger than confidence in Test 2 for repeated (*M* = 1.70, *SE* = .09) and Test 2 baseline conditions (*M* = 1.69, *SE* = .08). However, a one-way repeated measures ANOVA with Test 1, Test 2 repeated and Test 2 baseline conditions was not significant (*F*(1.11,58.58) = 0.87, *p* = .36, η^2p = .02) suggesting no statistical differences in confidence between conditions. These results conflict with data from Experiment 3a, where confidence was larger for Test 1 and Test 2 repeated conditions, compared to the Test 2 baseline condition. Such difference may be a reflection of the continuous confidence scale used in Experiment 3b.

Accuracy and confidence of Test 1 recognition modulates subsequent accuracy and updating

The second and focal aim of Experiment 3b Tested whether an interaction between recognition accuracy and confidence during Test 1 modulated face memory updating in Test 2. As mentioned, measures of face memory updating were calculated for trials of Test 1 correct high confidence and Test 1 correct low confidence conditions and Test 1 incorrect high and Test 1 incorrect low confidence. It was reasoned that bias measures should be more positive for high versus low confidence conditions, for both correct and incorrect Test 1 trials, supporting the prediction the subjective retrieval processes enhanced retrieval-induced updating of face memories. Exclusion criteria for analysis were applied to ensure that participants had sufficient trial numbers for each of the Test 1 correct high, correct low, incorrect high and incorrect low conditions, in order to calculate reliable proportion bias measures. Participants with fewer than 5 trials in any condition were removed (i.e. participants who had too few correct or incorrect trials), leaving an analysis sample of 51 participants[4.10]

Proportion bias. Firstly, a 2 (Test 1 accuracy; correct, incorrect) x 2 (Test 1 confidence, high, low) repeated measures ANOVA compared Test 2 proportion bias scores (see Figure 3.7 for descriptive statistics). The ANOVA showed main effects of Test 1 accuracy (*F*(1,50) = 43.96, *p* <.001, η^2p = 0.47) and Test 1 confidence (*F*(1,50) = 13.68, *p* <.001, η^2p = 0.22). The interaction, however, was not significant (*F*(1,50) = 0.96, *p* = .33, η^2p = .02). Planned comparisons showed that proportion bias was significant higher for Test 1 correct high versus correct low confidence conditions and Test 1 incorrect high

---

[4] A further exclusion criterion was applied to identify if participants had test 1 accuracy scores of 0.2, however all 51 participants passed this criterion.

versus incorrect low conditions. Table 3.2 also demonstrates that, for the correct high-low contrast, Bayes factors showed substantially more evidence in favour of the alternative versus null hypothesis, however the Bayes Factor for the incorrect high-low contrast was anecdotal.

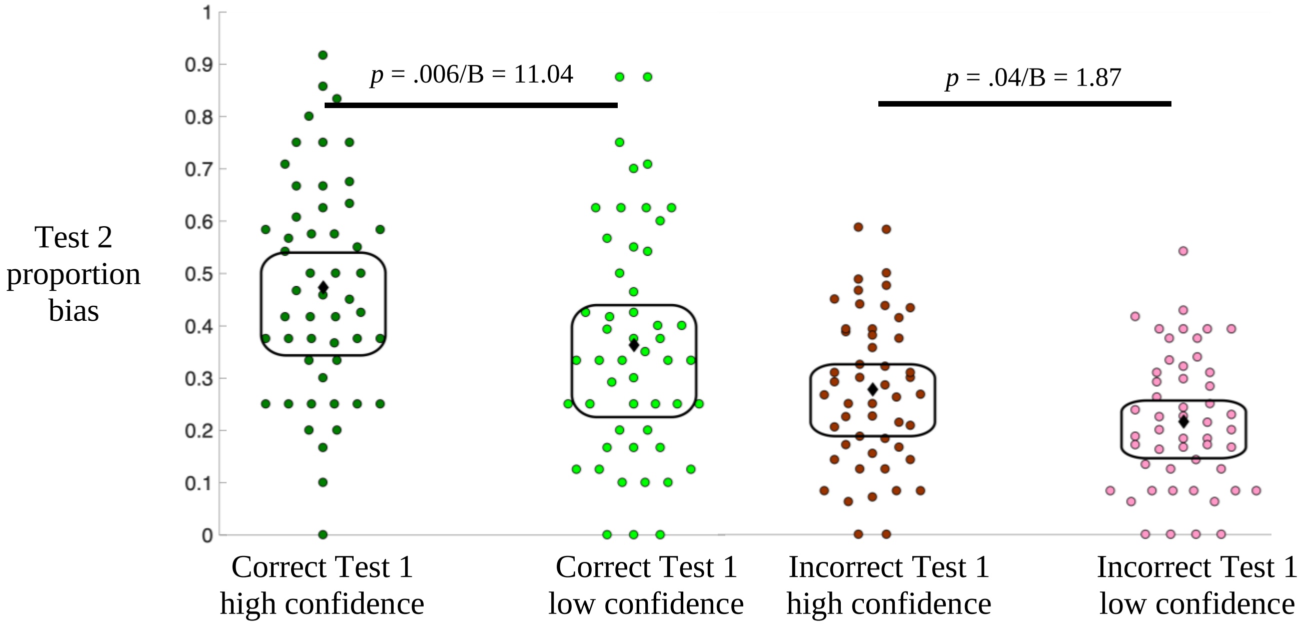

Figure 3.7. Proportion bias for correct and incorrect high and low confidence conditions.

Table 3.2. Inferential statistics for bias analysis in Experiment 3b.

High-low confidence comparison	df	*t*	*p*	*g*	BF	Prior
Correct, proportion bias	50	2.87	.006	0.40	11.04	0.44
Incorrect, proportion bias	50	2.13	.04	0.30	1.45	0.69

Relationship between Test 1 accuracy and Test 2 bias. The final analyses aimed to replicate an interesting relationship from Experiment 3a, whereby participants were more likely to repeat the same/similar recognition during Test 2 when the prior Test 1 responses was similar to the target face. To consolidate this finding, the Euclidean distance between the face selected at Test 1 and the target face for each trial (Test 1 error) was extracted within participants. In addition, for each trial the Euclidean distance between the face selected at Test 1 and the face selected at Test 2 was acquired (Test 2 bias). For all 40 repeated trials, a Pearson's correlation was then conducted between these two variables, with the mean correlation coefficient calculated across participants. The mean correlation between Test 1 error and Test 2 recognition bias ($r = .21$) was significantly higher than zero ($t(53) = 8.65$, $p <.001$, $g = 1.17$). In addition, the Pearson correlation between Test 1 error and Test 2 bias was re-calculated for Test 1 incorrect trials only, with the mean of these correlation coefficients ($r = .26$) also being significantly greater than zero ($t(53) = 7.73$, $p <.001$, $g = 1.05$). Furthermore, the Bayes Factors for these t-tests provided large evidence in favour of the alternative versus null hypothesis (Test 1 all trials; $BF_{+0(1.95,0.707)} = 1.73e +9$); Test 1 incorrect trials – $BF_{+0(1.83,0.707)} = 5.68e +7$). These results validate the finding that Test 1 recognition responses that were similar to the target face memory were more likely to be re-selected at later recognition, replicating the results from Experiment 3a. Figure 3.8 illustrates the consistency of the positive relationship between Test 1 accuracy and Test 2 bias across participants.

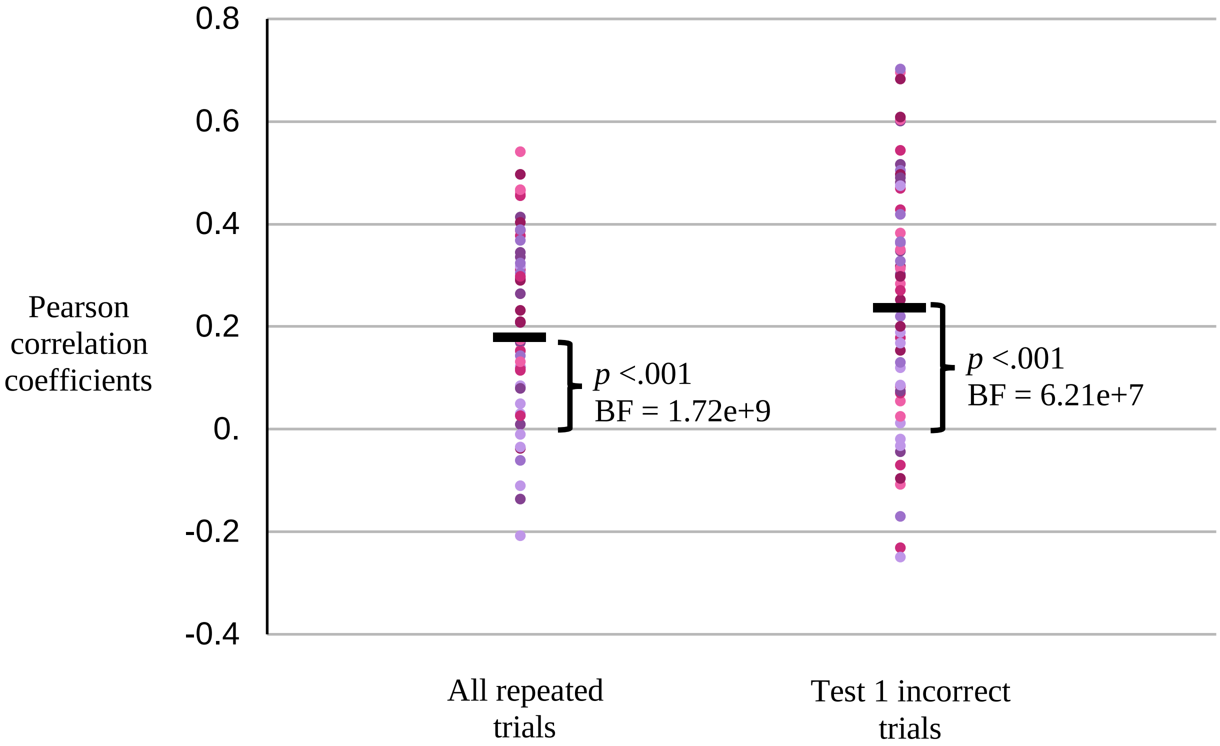

Figure 3.8. Correlation between Test 1 accuracy and Test 2 bias. Pearson correlation coefficients between Test 1 recognition error and Test 2 bias for all repeated trials and incorrect Test 1 trials. Mean correlation coefficients (horizontal black bars) suggest that Test 2 recognition was similar to Test 1 recognition responses when initial recognition was more similar to targets.

3.2.3. Discussion

The aim of Experiment 3b was to replicate the findings from Experiment 3a showing that; 1) objective and subjective retrieval varies across repeated and delayed retrieval, and 2) retrieval-induced updating of face memories is modulated by the accuracy and confidence of initial recognition. Using a modified version of the repeated recognition paradigm used in Experiment 3a, results from Experiment 3b replicated the second of these findings whereby participants were biased towards previous recognition responses when the initial recognition decisions was made with higher rather than low confidence. The same pattern of results emerged for both Test 1 correct and incorrect trials, suggesting that recognition confidence that is both accurate and erroneous can have positive and negative consequences on subsequent face memory retrieval (St. Jacques et al., 2013; St. Jacques and Schacter, 2013). In addition, results from Experiment 3b demonstrated that the size of recognition errors made during Test 1 (i.e. Test 1 error) correlated with the similarity of Test 1-Test 2 recognition responses (i.e. Test 2 bias). In other words, participants were more likely to make similar recognition responses during Test 2 as during Test 1 when responses in Test 1 were similar to the original target face. The replicability of these results from Experiment 3a to 3b further supports the validity of this relationship.

A final consideration of results from Experiment 3b analysed how retrieval accuracy and confidence varied across trial types. From Experiment 3b, it was shown that recognition accuracy during Test 1 was significantly larger than recognition accuracy in Test 2 for repeated and baseline trials. In addition, confidence for the Test 1 condition was larger than confidence for both Test 2 conditions (though not statistically significant). These results differed to those seen in Experiment 3a, where accuracy was lower for the Test 2 repeated versus Test 1, yet confidence did not differ between these conditions. The difference in results may represent a consequence of the continuous scale used in Experiment 3b, such that participants may have found it more difficult to use the continuous confidence scale compared to the categorical labelled scale in Experiment 3a, which could have led to noisier confidence measurements. This account can also explain why the bias difference effects were weaker in Experiment 3b compared to Experiment 3a, as indicated by smaller effect sizes for the correct and incorrect bias analyses.

3.3. General discussion

As mentioned previously, the results from the experiments in Chapter 3 are consistent with work examining the influence objective and subjective retrieval processes on episodic memory updating (St. Jacques et al., 2013; St. Jacques & Schacter, 2013). These studies established that enhanced reliving of autobiographical memories (a measure of subjective retrieval processes) increased the likelihood of recognising photographs of the events in a subsequent recognition task. What's more, increased reliving of memories in response to target photographs also increased the likelihood that lure photographs (presented after the target photograph) would be falsely recognised during the subsequent recognition test. Thus, enhanced subjective retrieval processes during an initial recognition task predicted the likelihood of memory updating as evidence by subsequent recognition.

The experiments from the present chapter build on the literature in several ways. First, the research in the present chapter provides evidence of updating of simpler episodic memories than has been studied before. That is, previous work concerning retrieval-induced updating has focused on memory updating of elaborate and relational visuospatial (Bridge & Paller, 2012; Bridge & Voss, 2014) or autobiographical memories (St. Jacques et al., 2013; St. Jacques & Schacter, 2013). One of the questions posited at the beginning of this chapter queried whether retrieval-induced updating occurs during simpler forms of episodic memory retrieval, such as the recognition of item memories. The data from these experiments provides novel evidence that recognition-based retrieval processes can induce memory modifications for item memories, and that such updating effects are modulated by recognition confidence during the initial retrieval attempts.

In addition, the present experiments show that a different aspect of subjective retrieval, that being confidence in retrieval decisions, was associated with enhanced face memory updating. An important distinction between the present study and previous work is that reliving judgements in research by St. Jacques et al. (2013) and St. Jacques and Schacter (2013) were specifically related to target memories only, and that updating of memories with foil information resulted from presenting misinformation following a retrieved memory. In contrast, the present experiments show that updating of memories occurred when confidence judgements were made with regards to recognition of selected face images, in that participants provided confidence judgements for both target memories (if correct) and distractor faces (if making errors), and that these high confidence responses predicted repetitions of accurate and erroneous recognition responses. Such distinctions are important to emphasise when considering the underlying explanations of the retrieval-induced updating of face memories.

One proposed mechanism to explain increased bias during Test 2 following high rather than low confidence retrieval during Test 1 considers that enhanced subjective retrieval processes were more likely to reactivate face memories during Test 1 that predicted enhanced bias in Test 2 (e.g. Johnson et al., 2009; Leiker & Johnson, 2014), consistent with the view that reactivation is critical for memory updating (Hupbach et al., 2007; 2008; 2009; Sederberg et al., 2011; St. Jacques et al., 2013). It should be emphasised that distractor faces had never been seen prior to the first test, therefore how could these memories have a neural representation that would be reactivated during Test 1. As proposed, sub-optimal encoding of target faces during learning may led to a distorted, or ‘fuzzy-trace’ representation of these memories (e.g. Brainerd & Reyna, 2002). Alternatively, taking assumptions from face space models (Lewis, 2004; Valentine, 1991; Valentine et al., 2015), distractor faces may have been represented within an error boundary of target images in face space. Both of these suggestions can potentially explain how distractor faces would be reactivated during Test 1 as a function of enhanced recognition confidence.

The finding that recognition confidence, for both correct and incorrect face recognition decisions, predicts face memory accuracy/updating could also be explained by recent findings suggesting that encoding processes are engaged during retrieval. It is possible that high recognition confidence during Test 1 increased encoding of recognition choices, possibly mediated by enhanced attentional resources to face memories that evoked high confidence signals (Ciaramelli et al., 2008), in contrast to low confidence recognition decisions. When high confidence is made with correct Test 1 recognition attempts, the encoding of these face memories is enhanced, whereas low confidence decisions, despite incorporating some re-learning of the target face, are encoded less due to the low confidence choice. Similarly, high confidence for incorrect Test 1 recognition choices could increase the encoding of these retrieval errors, leading to biased face recognition memory during repeated retrieval. In contrast, low retrieval confidence may fail to engage encoding of incorrect retrieval decisions, reducing the likelihood of face memory updating during subsequent recognition.

However, it has to be acknowledged that the current research design only enabled an investigation of correlational patterns between confidence and subsequent bias, and therefore a causal relationship cannot be confidently inferred. It is also possible that confidence and bias correlated due to some other factor, such as the strength of initial face memory encoding. This might especially be the case for correct trials, where high confident accurate responses might have been repeated across tests because some faces were very strongly and accurately encoded, whereas low confidence accurate responses might have been “lucky guesses”. The use of ERP measures with the present face recognition paradigm (see Chapter 5) could help determine which of these proposed accounts is correct. For example, and consistent with previous ERP studies of retrieval-induced updating (Bridge & Paller, 2012; Liu et al., 2018), ERPs during Test 1 associated with accurate or confident responses could reflect reactivation of memories. In contrast, different ERP activity during Test 1 could be visible for high confidence responses that are repeated in Test 2, consistent with an ‘encoding-during-retrieval’ account.

An interesting finding from the present studies was that face memory updating was more evident when faces selected during Test 1 recognition decisions had greater resemblance to the original target face memory. That is, using the continuous measures of memory accuracy (Euclidean distance between face images), recognition errors with smaller Euclidean values during Test 1 were correlated with enhanced recognition bias (similar face choices) during Test 2 for the majority of participants. A possible reason for these findings can be found in terms of face space model assumptions postulating that all faces have a degree of error surrounding the representation location (Valentine et al., 2015), particularly following poorer face encoding (Valentine, 1991). Distractor face representations that have not been encoded or seen previously, but are encompassed by this location error will be more perceptually similar to an encoded target face. It is possible that distractors that bore greater resemblance to the original target memory (therefore lying within the target location error “sphere”) were more likely to be selected during Test 1 due to a higher level of activation in the face space (Lewis, 2004), which could adjust and update the location of a face representation in face space when these similar but erroneous representations are recognised to be the target face.

With this suggestion, it could be argued that targets and similar distractors were ‘merged’ together so that the target face representation became ‘morphed’ with the similar distractor image. In comparison, distractor faces that are perceptually dissimilar to the target face, would be encoded at a more distal location in face space when presented during Test 1, beyond the target face error boundary and thus will have weaker representational strengths. In this scenario, target and distractor memory representations would not merge and would therefore be in competition for selection during Test 2 recognition. Given the present data, such competition would more likely lead to faces being selected during Test 2 that resembled the correct target face memory (given the elaborate encoding procedure of targets during learning) rather than selection of the dissimilar distractor faces (given the short opportunity for encoding during Test 1 recognition tasks).

These arguments can be supported by recent evidence showing that encountering similar information across repeated events tends to merge/blur that information in memory (Li et al., 2019), where blurred memories become more coarse-grained representations that reduce the precision of retrieval attempts (Sun et al., 2017). On the other hand, encountering less similar information produces competition between separate memory traces at subsequent retrieval attempts. This argument may be relevant for some participants who showed negative correlation patterns between test 1 error and test 2 bias (i.e. selecting similar faces across tests when initial recognition was more dissimilar to the target). Therefore, it may be the case that multiple mechanisms can lead to memory updating, an issue addressed in the general discussion of this thesis (see Chapter 7). In the Li et al. (2019) and Sun et al. (2017) studies, such retrieval competition led to dissimilar lure information being selected more so than original encoded information, attributed to an ‘erasure’ of the original memory. This may have also been the case in the present experiments, however there would be a tendency that retrieval competition favoured the selection of targets versus dissimilar distractors owing to the strength of target representations following learning and prior testing.

These ideas can be linked to the neurocomputational mechanism of pattern separation, whereby the hippocampus assigns unique representations to individual events that do not overlap with other, similar memories (Leal & Yassa, 2018; Yassa & Stark, 2010). For dissimilar distractors, it is possible that unique representations were formed during Test 1 that did not overlap with the target memory representation, leading to retrieval competition in Test 2. Nevertheless, these arguments are purely speculative based on behavioural data alone. Advanced neuroimaging analyses could also be used in future work, such as representational similarity analysis (Kreigeskorte et al., 2007) and pattern-based analyses (Lee & Kuhl, 2016; Kuhl & Chun, 2014), to assess the representational activation of face memories across multiple learning and retrieval attempts, according to the size of retrieval errors made during initial face recognition.

The present study demonstrated the consequences of repeated retrieval on both objective and subjective retrieval processes. Specifically, objective face recognition accuracy was shown to be highest during the initial Test 1 recognition task, compared to Test 2 recognition accuracy for face memories that underwent repeated (repeated condition) or delayed (baseline condition) recognition attempts. However, the subjective experience of retrieval across the face recognition tasks did not necessarily match objective face recognition performance. For example, participants were shown to have higher recognition confidence during Test 2 recognition for face memories undergoing repeated retrieval, compared to face memories in the baseline condition. Furthermore, Experiment 3a demonstrated that participants reported more ‘multiple memory’ experiences following repeated face memory retrieval, as opposed to reporting an ‘absence’ of face memories for memories in the baseline condition. These findings converge with results from the studies by St. Jacques and colleagues (St. Jacques & Schacter, 2013), who showed that recognition confidence for lure photographs during the second recognition task was higher for photographs that were shown during the initial recognition task. Results from the present experiments are also similar to evidence showing that confidence for incorrect recognition is increased when stimuli are semantically related to original memories (DeSoto & Roediger, 2014; Roediger & DeSoto, 2014). The present data also corresponds to findings in applied research showing that eyewitnesses are just as confident in incorrect face recognition decisions across repeated retrieval attempts, compared to correct face identifications that did not undergo repeated retrieval (Goodsell et al., 2009). Overall, these findings converge with previous work emphasising dissociations between objective retrieval accuracy and subjective retrieval confidence.

Such dissociations between objective and subjective retrieval processes, and how these different mechanisms are modulated by repeated recognition, raises a relevant question as to how these different memory processes interact during retrieval and updating. Plenty of research implies that objective and subjective processes do not correlate (DeSoto & Roediger, 2014; Roediger & DeSoto, 2014; St. Jacques & Schacter, 2013) evidenced in the present chapter where participants, for some trials, report high confidence during false recognition attempts. Furthermore, such high confidence errors were related to memory distortions more so than when distractors were recognised with low confidence. As mentioned, the neurocognitive mechanisms by which subjective retrieval processes could enhance updating requires further study, an issue addressed in the current thesis in Chapters 5 and 6.

In conclusion, the present experiments suggest that retrieval-induced updating of face memories is modulated by both objective accuracy and subjective retrieval during initial retrieval attempts. The research in this chapter extends on previous findings by suggesting that item memories are susceptible to updating via recognition attempts. What the present set of experiments do not answer, however, is whether active retrieval attempts are critical for the observed updating effects. Consequently, the aim of the research in the next chapter (Chapter 4) was to determine whether face memory updating is critically related to the act of retrieval or is instead due to presenting of misinformation between encoding and retrieval of face memories.

Chapter 4 - Comparing retrieval versus re-study on face memory updating

The results from Experiments 3a and 3b suggest that retrieval-induced updating of face memories is enhanced when initial recognition decisions are made with high rather than low recognition confidence, regardless of the accuracy of the decision. These experiments, however, did not address whether retrieval during test 1 was critical to induce memory updating. Instead, it is plausible that the mere perception and encoding of distractor faces during test 1 could have biased subsequent responses towards those distractors, without the need for memories to be reactivated during test 1. The experiments in Chapter 4 were conducted to test whether the response biases that emerged after testing in the previous studies were caused by active retrieval attempts, or were instead simply due to the presentation of distracting information between learning and final recognition tests.

As indicated from the literature review in the present thesis, retrieval is known to be a powerful learning mechanism that can induce episodic memory updating. The 'retrieval practice' literature suggests that retrieval of memories, compared to re-studying learnt information, is more likely to lead to repeated responses during a subsequent criterial retrieval task (Roediger & Butler, 2011). The retrieval practice effect has been demonstrated using a variety of retrieval task formats, including recognition memory by using task formats such as multiple choice questions (MCQ; Butler et al., 2006; Marsh et al., 2007; Roediger & Marsh, 2005), where participants are presented with multiple response options (one correct and several incorrect responses) to a given question. Using this task, several studies have shown that taking an MCQ can have opposing effects on memory updating. When participants correctly retrieve information during an initial MCQ, then accuracy on the criterial test is enhanced compared to when correct information is re-studied during initial MCQ. The same pattern of findings is shown when incorrect information is falsely recognised during initial MCQ’s, showing the detrimental impact of MCQ testing when initial retrieval is erroneous (Marsh et al., 2007; Roediger & Marsh, 2005).

Despite convincing evidence showing retrieval practice effects, it should be noted that retrieval practice benefits are not always apparent. For example, retrieval practice benefits have sometimes been shown only following longer delays between initial and criterial testing, with no retrieval practice benefit observed at shorter time intervals of 5 minutes between initial and criterial tasks (Rowland & Delosh, 2014; Smith et al., 2013; Wheeler et al., 2003). What’s more, several studies have observed retrieval practice benefits only if accuracy during initial retrieval is high (Rowland, & Delosh, 2014; Smith et al., 2013). However, the difference in accuracy rates between retrieval and re-study conditions highlights a limitation of retrieval practice studies, in that retrieval and re-study conditions may vary in the proportion of trials where correct information is attended. For example, retrieval accuracy will vary between participants who have better/poorer memory for prior events. However, in a standard re-study condition, the correct answer is almost always reinforced as participants are told to re-study the correct answer. In this case, participants are likely to have better accuracy during criterial testing following re-study versus retrieval tasks. Thus, retrieval practice effects may not be due to processing differences between retrieval/re-study tasks, but could instead be due to confounds in terms of how much the retrieval and re-study conditions encourage encoding of accurate information.

Addressing this issue, Bridge and Voss (2014) developed a learning – refresh – recognition paradigm, where the degree of retrieval accuracy/error of object-location associations during the refresh task was matched between an active retrieval and passive re-study task, an improvement on prior research using retrieval vs. re-study manipulations which had not matched the amount of error that was reinforced by the conditions. In their paradigm, active retrieval blocks preceded passive re-study blocks so that the items in the re-study block could be matched to locations that participants retrieved during the active retrieval task in the previous block. Even with the degree of accuracy/error controlled for in this way, Bridge and Voss (2014) showed that memory for object-locations was more likely to be updated following active retrieval versus passive re-study conditions. Building on the findings from Bridge and Voss (2014), the first experiment of Chapter 4 aimed to compare the influence of retrieval (via recognition) versus re-study on the updating of face memories. This paradigm builds on previous work using MCQ formats with the retrieval-practice effect (Marsh et al., 2007; Roediger & Marsh, 2005). However, Experiment 4a adopted a procedure to match recognition accuracy during active retrieval and passive re-study conditions during refresh (Bridge & Voss, 2014) by using face stimuli developed from face space models of face recognition (Solomon et al., 2015; see section 4.1.1).

4.1. Experiment 4a

The aim of Experiment 4a was to compare the influence of active retrieval versus passive re-study on the updating of face memories. A paradigm was developed where participants completed two cycles of tasks, both incorporating learning, refresh and final recognition tasks. In both blocks, participants learnt 30 targets during learning before completing either an ‘active’ retrieval or ‘passive’ re-study refresh task on 20 of the 30 face targets. For both retrieval and re-study tasks, one target face was shown with four distractor faces related but perceptually distinct from the target face. During retrieval, participants were asked to select which of the five faces was shown during learning. In contrast, the re-study task required participants to encode one of the five faces for the subsequent recognition task, with the particular face indicated for encoding determined by their response on a matched trial from the previous active retrieval task. That is, pairs of trials were matched across retrieval and restudy tasks in terms of how similar targets and distractors were to each other (as measured by the Euclidean distance between them in face space), even though the actual face images (and locations in face space) were different between the two trials. If participants made a correct response on the retrieval trial, the corresponding re-study trial asked them to encode the correct response. If participants made an error on the retrieval trial, then they were asked to encode a distractor on the corresponding re-study trial which matched the degree of error to the retrieval trial (i.e. with the same target-distractor Euclidean distance in face space). Following refresh, a final recognition task was completed which had a similar design to the active retrieval task, but was completed for all 30 face targets (20 repeated trials previously shown during refresh, 10 baseline trials not shown during refresh).

From this paradigm, analysis was focused on the patterns of recognition accuracy and updating during the final recognition task, compared following retrieval (in block one) and re-study (in block two). For final recognition accuracy, the proportion of targets that participants correctly recognised was calculated separately for repeated trials and baseline trials. In addition, final recognition bias was used as a measure of memory updating, to assess the extent to which participants were biased by selected/encoded distractors from refresh during final recognition. This bias measure thus enabled a comparison between the amount of face memory updating following active retrieval versus passive re-study.

Consistent with the prior literature, it was hypothesised that retrieval compared to re-study during refresh may have a stronger effect on final recognition measures. However, given the short delay between refresh and final recognition tasks, it was possible that no retrieval practice benefits would be found (Rowland & Delosh, 2014; Smith et al., 2013; Wheeler et al., 2003). If a retrieval practice benefit was present, then final recognition accuracy should be larger following active versus passive refresh. This prediction would only be expected for repeated trials, with no difference expected for baseline trials as these memories should be minimally influenced by the refresh manipulation. In addition, it was hypothesised that retrieval during refresh would lead to larger final recognition bias, compared to re-study. These predictions are consistent with the view of retrieval as a powerful mechanism for learning (Bridge & Voss, 2014; Roediger & Butler, 2011).

4.1.1. Method

Participants

Eighty-one undergraduate participants ranging in age from 18-28, with 68 females (M_{age} = 19.46, SD_{age} = 1.78) were recruited from the University of Kent. This sample size was recruited to maximise participant numbers during a limited recruitment time window. However with this sample size, there was 99% power to detect a medium effect size (0.5) and 71% power to detect a small effect size (0.3) with a two-tailed alpha of .05. Participants received course credits in return for participation, and the study had ethical approval received from the University of Kent School of Psychology ethics board. Participants were required to have normal/corrected to normal vision and be aged between 18-35. Nine participants were removed due to failures of task compliance during the experiment (sleeping, not looking at computer screen, etc).

Materials

Face stimuli. Artificial face space models were used to create face stimuli (Solomon, Gibson & Mist, 2015). Face stimuli were sampled from a face space constructed of 29 components. From this face space, 70 face sets were created with each set containing five faces (images 1-5). Initially, 70 face locations were sampled at a constant radius from the face space origin, with target to target Euclidean distances ranging from ~5,000 to

~15,000 (see Figure 4.1). For the experiment, 10 face sets were removed due to image artefacts in these sets.

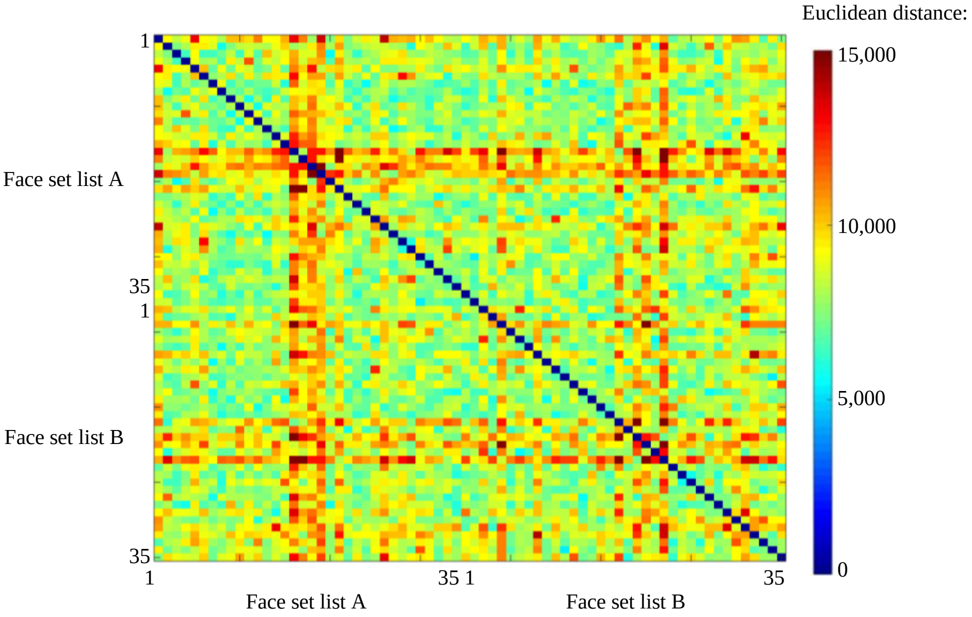

Figure 4.1. Euclidean distance between initial face locations used in Experiments 4a-c. Colour map representing the Euclidean distances between the 70 initial face locations (for all 70 face sets) in an artificial face space. The range of distances from 5,000-15,000 ensures sufficient perceptual dissimilarity between face sets.

For all 60 initial face locations, five additional face locations were randomly sampled from a uniform distribution limited to 1000-6000 units around each face location. Thirty face sets were allocated to list A, with 30 ‘matched’ sets allocated to list B. Face set lists were matched to equal the pairwise Euclidean distances between nine images within matched face sets. For example, Figure 4.2 shows the pairwise Euclidean distances between images 1-5 for set one in list A and set one in list B. Figure 4.2 demonstrates that sets one in list A and B have the exact same pairwise Euclidean distances between the five images within each set. Of further importance, the faces from list A were perceptually dissimilar from the matched faces in list B. This matching procedure removed any confounding influence of face stimuli variances in producing differences in face recognition measures across block.

For the face recognition task, one of the five faces within a set was designated as the experimental target face shown during learning, refresh and final recognition tasks. Each of the five images across sets was selected as the experiment target and equal number of times across participants (i.e. target assignment was counterbalanced). The remaining four faces in each set were ‘distractor’ faces that were presented during only the refresh and final recognition tasks. In addition, face stimuli were counter-balanced across participants according to whether face sets were presented during refresh and final recognition (repeated trials) or only presented during final recognition (baseline trials). Furthermore, face set lists were counter-balanced across participants to be presented in block one or two.

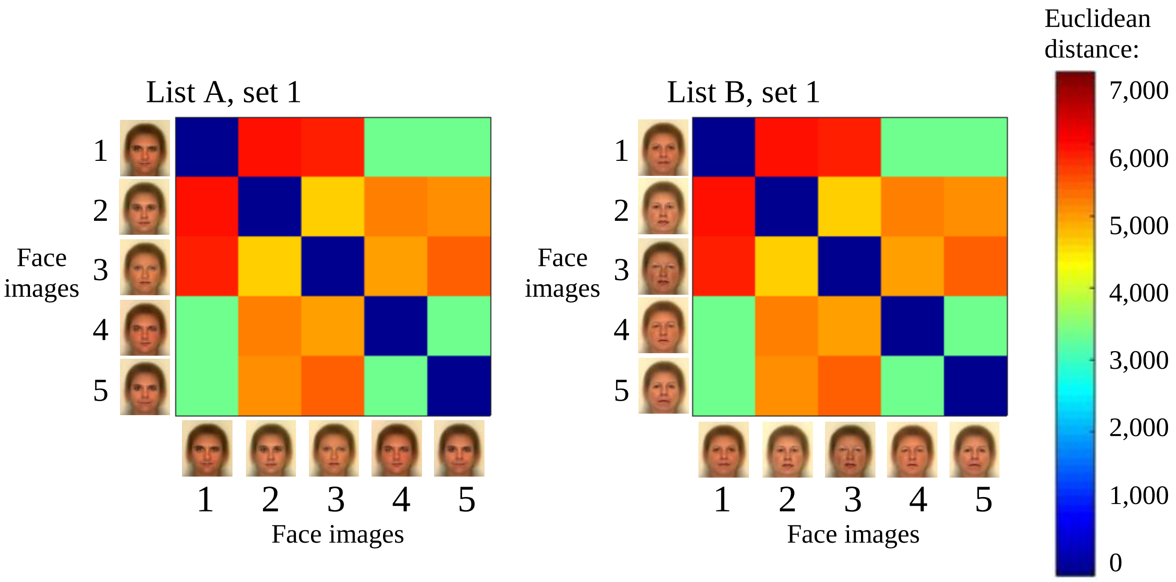

Figure 4.2. Euclidean distance between faces within sets. Colour maps shown for face images in face set 1 (list A top left, list B bottom left). Face sets were matched according to the pairwise Euclidean distances between the 5 face images within the pairs of sets.

Apparatus

The experiment was completed on a Dell optiplex 9020 desktop computer with PsychoPy (Peirce, 2009). The screen measured 51cm x 28.4cm, with a resolution of 1920 x 1080 pixels. All participants completed the experiment on the same computer in a Psychology lab room.

Procedure and design

Participants completed two blocks, with both blocks containing a similar task procedure of learning, filler, refresh, filler and final recognition tasks. The only difference between blocks was the type of task completed during refresh. Block one refresh was an ‘active’ retrieval task (see Figure 4.3), whereas block two refresh contained a ‘passive’ re-study task (see Figure 4.4). This within-subjects design therefore manipulated the refresh task (active retrieval, passive re-study), as well as the type of trial during final recognition (repeated, baseline).

Learning. Participants completed three cycles of face learning tasks, with all 30 face targets shown in all three cycles. For each cycle, face targets were presented one at a time for 5000ms each in a randomised order, preceded by a 500ms fixation. All face pictures subtended a 5.12 x 5.88 visual angle when participants sat at a distance of 75cm from the screen. During face presentation, participants were instructed to memorise face images for later memory tasks. Participants were asked to make a rating for each face target in order to facilitate encoding. During cycle one, participants decided whether the person of each face was trustworthy (key press 5) or not trustworthy (key press 1). Cycle two involved rating faces as attractive (key press 5) or unattractive (key press 1). Finally, the third cycle required participants to indicate whether a face was feminine (key press 5) or masculine (key press 1). The traits used in the rating task were informed by Oosterhoff and Todorov (2008) who showed that these different trait ratings tend to be based on different aspects of face images. Participants were told of each rating task prior to beginning each cycle and made their responses whilst the face was on screen.

Letter search filler task. Following learning, participants completed Treisman and Gelade’s (1980) visual letter search task as a filler task. The purpose of this task was to separate learning and refresh tasks so that performance during refresh task, particularly in block one, was due to long-term retrieval as opposed to short-term working memory. Seventy-two letter arrays, containing a mixture of letter type (X, O & N), frequency (1, 5, 15 & 30) and colour (red, green & blue), were presented one at a time. Participants were tasked with searching for a blue letter which was present in 18/72 pictures. Participants indicated whether the array had a blue letter (key press 5) or did not have a blue letter (key press 1). Trials were self-paced, with all 72 trials completed in around 1 minute.

Retrieval during refresh (block one). The refresh task completed in block one was a face recognition task. This task presented 20/30 of the face sets, with one set shown per trial. For each trial, a target face was presented with the four distractor faces from the corresponding face set. These five faces were presented sequentially, numbered from 1-5 and shown for 1250ms each, with a 1000ms fixation preceding each face. All face images subtended a visual angle of 5.12 x 5.88. Across trials, each face image was located in each of the five positions an equal number of times. After viewing all five faces, participants were prompted with a ‘?’on screen, which cued participants to press 1-5 on the keyboard to indicate the position of the face they recognised to be the target. After responding, the next recognition trial began, with face set order randomised for each participant. Because this experiment was designed to compare active retrieval with re-study, it was not possible to include confidence ratings in the task (completing such ratings would only have made sense for participants in the retrieval task but not in the re-study task, and including confidence ratings only in the retrieval task would have introduced confounds).

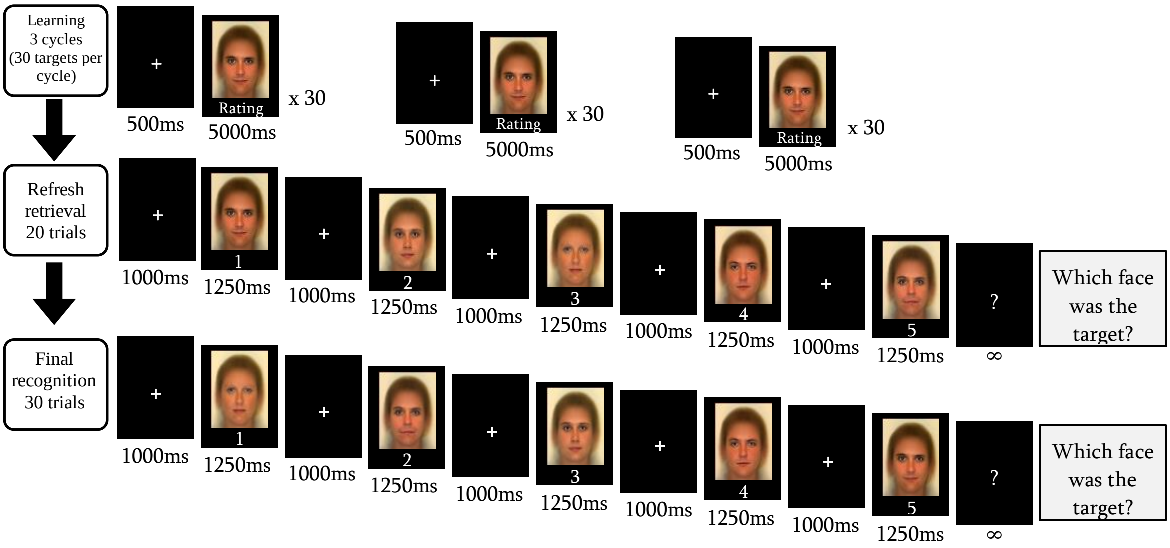

Figure 4.3. Block one procedure for Experiment 4a. During learning, participants encoded 30 targets, with all targets presented in three cycles. Next, participants completed 20 ‘active’ face recognition trials where participants were asked to select a target face shown amongst four distractor faces. The same face recognition task was completed during final recognition, but for all 30 face targets (20 repeated trials, 10 baseline trials).

Re-study during refresh (block two). The refresh task in the second block was a face re-study task. Face sets used in the second block were the matched face sets taken from the face set list that was not shown during block one (e.g. if list A was used in block one, list B was used in block two). The re-study task used a similar trial format as the retrieval during refresh task to avoid confounds between refresh tasks, with the five faces from a set presented sequentially, numbered from 1-5 and shown for 1250ms each, with a 1000ms fixation preceding each face. Four of the five fixation crosses that were presented before face images were coloured grey, with one fixation cross presented in white. The white fixation cross was used as a cue for participants to selectively encode the face that immediately followed the white cross. Importantly, the position of the white fixation cross was determined according to participants responses during the refresh task in block one. That is, if participants selected the first image of set one during the retrieval refresh task, then for set one shown during re-study, the first image would be preceded by a white fixation cross (see Figure 4.4). Critically, the faces matched by position were also matched according to Euclidean distance, ensuring that the faces participants selected during passive re-study were equal in error to faces selected in active retrieval tasks. This procedure was done for all refresh trials in order to match the Euclidean distance between target faces and faces selected during refresh for both retrieval and re-study tasks, meaning that the re-study task was always completed in block two (similar to Bridge & Voss, 2014). After seeing all five faces, participants pressed the key associated with the position of the face they were asked to encode. Critically, participants were instructed to focus on learning the cued face to be remembered for the subsequent recognition task, ensuring that participants were paying attention to face images in the re-study condition.

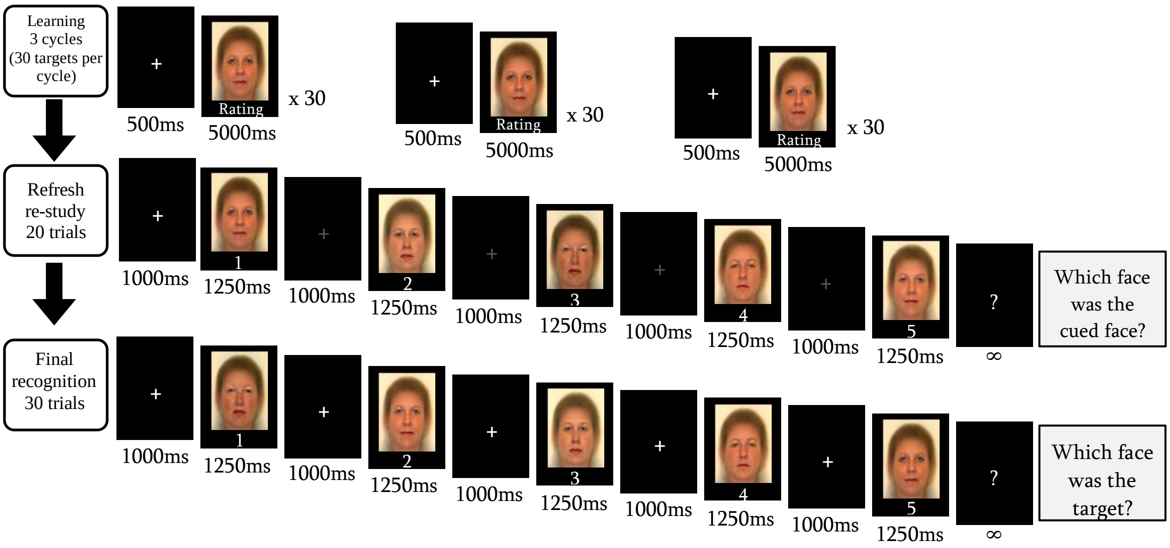

Figure 4.4. Block two procedure for Experiment 4a. During learning, participants encoded 30 targets, with all targets presented in three cycles. Next, participants completed 20 'passive' re-study trials where participants were asked to encode one of five faces. During final recognition, participants were asked to select a target face amongst four distractors for trials shown during refresh (repeated condition) and trials not shown during refresh (baseline condition).

Letter search filler task. The same letter search task separated the end of the refresh recognition and start of final recognition. Letter array order was randomised and took around 1 minute to complete.

Final recognition. The final recognition task involved the same trial format as described in the retrieval refresh task, with 30 trials completed. Of these, 20 were 'repeated trials' that were shown during the previous refresh task, and 10 'baseline' trials that were not shown during refresh. For repeated trials, the order of the five faces within each trial was randomly reshuffled from the previous refresh trial. Importantly, participants were instructed to focus on recognising the face they encoded during the learning phase, and ignore the face that they retrieved/encoded in the previous refresh task. The 30 final recognition trials were presented in a random order. After all trials were completed in block one, participants received a short break for a few minutes before completing block two.

Data analysis

Performance during the final recognition task was analysed to compare the effects of retrieval versus re-study on face recognition accuracy and updating. All measures reported were calculated separately for each block. Refresh target selection was calculated as the proportion of refresh trials where the target face was selected. Final recognition target selections for repeated trials was calculated as the proportion of repeated trials where the target was correctly recognised during final recognition. In addition, for the 10 ‘baseline’ trials, final recognition target selections for baseline trials was calculated as the proportion of baseline trials where the target face was selected during final recognition. Of these measures, refresh target selection was necessarily identical across the two blocks, given the matching procedure used which fixed the responses in block two based on block one performance. However, final recognition target selection rates could have differed according to the accuracy of participants memory for targets during final recognition. This analysis was exploratory, with final recognition accuracy compared as a function of block (one vs. two) and refresh manipulation (repeated vs. baseline trials).

A final recognition bias measure was calculated to determine the extent to which participants were biased by prior refresh errors during final recognition. This analysis included only repeated trials where a distractor face was selected during the initial refresh phase (i.e. falsely recognised as a target by the participant during the retrieval block, or indicated for encoding during the re-study block). For these trials, a final recognition bias measure was calculated as the difference in proportion of trials where participants selected the same distractor as during refresh, versus when they selected one of the other distractors. For “prior error” repeated trials, the proportion of trials where the same distractor was selected during final recognition was calculated. Next, the proportion of “prior error” trials where participants selected a distractor face during final recognition that was different to the one selected during refresh was calculated. This number was divided by three in order to calculate the proportion of trials that participants selected one of the three possible non-bias choices during final recognition, thus estimating the likelihood of participants making non-biased errors. To provide a single measure of final recognition bias, this proportion of non-biased error responses was subtracted from the proportion of biased error responses to provide a final recognition bias difference score that estimated how much more likely participants were to make a biased as opposed to non-biased error on the final test. The current experiment used final recognition bias difference scores rather than just proportion bias (used in Experiments 3a and b) as proportion bias measures may be confounded by differences between refresh tasks (see Chapter 2, Section 2.2.1 for more details).

The final recognition bias difference score was used as the key measure of face memory updating; the more positive the bias difference score was, the more participants were biased towards repeating the same recognition error during final recognition. In contrast, if the difference score was no different to 0, then participants were equally likely to repeat an error as making a non-repeated error during final recognition. Finally, a more negative bias difference score signifies that participants were more likely to select a different distractor face during final recognition compared to repeating errors during final recognition. Hence, it was predicted that the bias difference score should be significantly more positive than zero, if participants were biased by the distractors they selected or were told to encode during refresh. If bias is enhanced by active retrieval attempts, then this bias score should be more positive following retrieval compared to re-study during refresh, consistent with the hypothesis that active retrieval attempts enhance encoding and thus induce face memory updating.

4.1.2. Results

Target selections during refresh and final recognition

Figure 4.5 shows target selection measures during refresh and final recognition, separately for blocks one (retrieval) and two (re-study). Proportion target selections in all conditions was significantly higher than chance performance of 0.2 at p <.001. Due to the matching procedure used in the passive refresh condition, proportion of targets picked during refresh was equivalent during retrieval and re-study. However, for target selection during final recognition, a 2 (trial; repeated, baseline) x 2 (block; one, two) repeated measures ANOVA was conducted, showing no main effect of trial ($F(1,71) = 0.33$, $p = .57$, $\eta^2p = .005$), but a significant main effect of block ($F(1,71) = 10.85$, $p = .002$, $\eta^2p = .13$). In addition, there was no significant interaction ($F(1,71) = .04$, $p = .85$, $\eta^2p < .001$). Planned paired samples t-tests (Bonferroni corrected $\alpha = .025$) showed that target selection rates was significantly higher in block one than block two for repeated trials ($t(71) = 3.07$, $p = .003$, $g = 0.36$), but also marginally significantly higher in block one versus two for baseline trials ($t(71) = 2.10$, $p = .04$, $g = 0.25$). Thus, the results did not indicate a specific retrieval-practice effect for repeated sets only, since baseline sets showed a similar pattern.

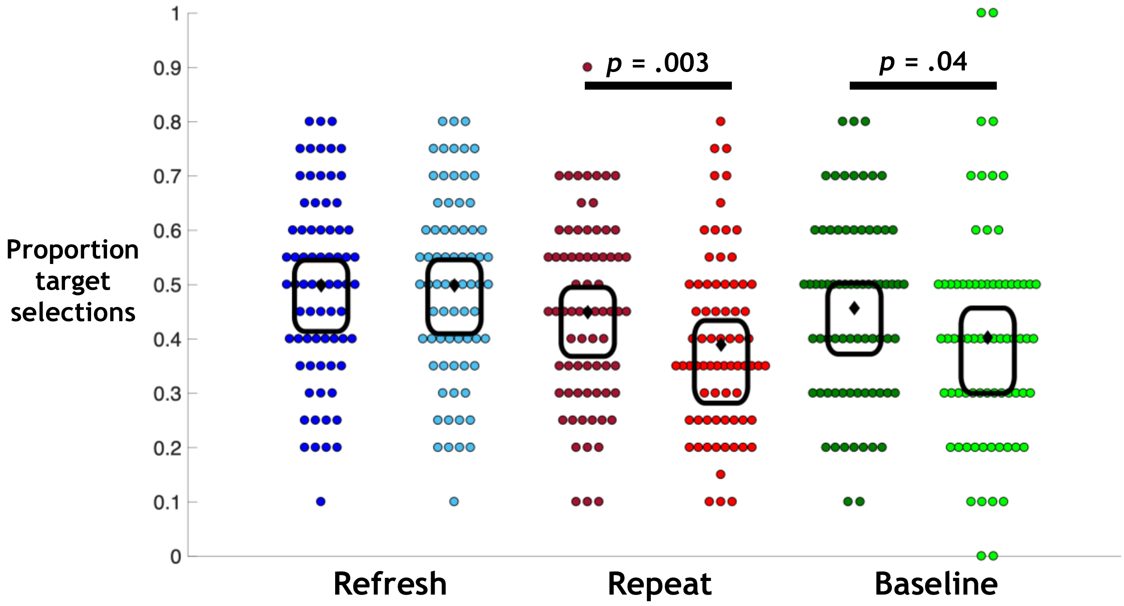

Figure 4.5. Statistics for target selection rates in Experiment 4a. Proportion target selections during refresh and final recognition (repeated and baseline trials) for block one (retrieval task, darker colours) and block two (re-study task, lighter colours). For both repeated and baseline trials, proportion target selections was significantly higher in block one versus block two.

Final recognition bias enhanced following retrieval versus re-study

The next analysis compared the effect of the refresh manipulation on the final recognition bias difference scores[1]. As a reminder, a more positive bias difference score indicates that participants were more likely to select a distractor that they had previously selected/been asked to encode during refresh, rather than switch between distractor recognition responses from refresh to final recognition. The final recognition bias difference score was significantly higher than 0 following both retrieval (*t*(71) = 6.16, *p*

---

[1] All 72 participants were included in the bias analysis. In an alternative analysis (not presented), the same analyses were repeated after removing 5 participants who had less than 0.2 accuracy during refresh retrieval (i.e. those who show poor memory for targets). However, the conclusions from both analyses are the same, therefore the whole sample is included to avoid introducing biases in the sample.

<.001, *g* = 0.73) and re-study refresh (*t*(71) = 3.17, *p* = .002, *g* = 0.37), confirming that both refresh manipulations showed evidence of bias, that may have been caused by encoding of face distractors during refresh. However, as seen in Figure 4.6, the final recognition bias difference was significantly higher following active retrieval versus passive re-study during refresh (*t*(71) = 2.80, *p* = .007, *g* = 0.33; see Figure 4.6).

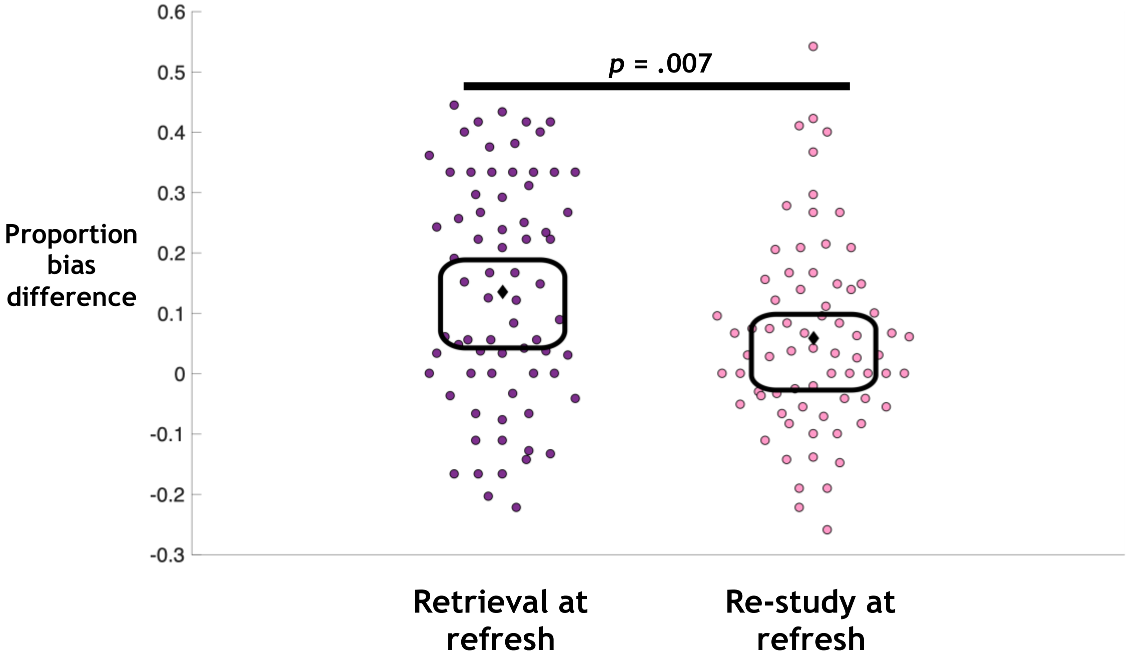

Figure 4.6. Final recognition bias difference measures in Experiment 4a. The bias difference was significantly higher following retrieval versus passive re-study at refresh.

4.1.3. Discussion

The results from Experiment 4a compared the influence of retrieval or re-studying face memories on the updating of these memories. The key results from this experiment showed that participants had improved accuracy for target faces during final recognition (for both repeated and baseline trials) when the preceding refresh task required retrieval rather than re-study. In addition, participant's memory during final recognition was biased more towards recognition errors made during the preceding refresh task involving active retrieval rather than errors they were instructed to encode during re-study. These results support the prediction that active retrieval attempts are a stronger inducer of face memory updating compared to re-study.

The results are supportive of the conclusion that retrieval is an important mechanism for the updating of face memories. Such evidence is consistent with literature of the retrieval practice effect via recognition-based testing (Marsh et al., 2007; Roediger & Marsh, 2005), as well as supporting neurocognitive studies comparing retrieval versus re-study on memory updating (Bridge & Voss, 2014). The Bridge and Voss (2014) study in particular had a similar design as Experiment 4a by including a re-study task where the information that participants were asked to encode in the re-study condition was matched (in terms of degrees of error) to participants' responses made in a preceding retrieval refresh task. This study design thus controls rates of accuracy between retrieval and re-study conditions by matching the number of errors that are at the focus of attention between refresh tasks. Furthermore, the degree of recognition errors between refresh tasks was matched, minimising potential confounding effects of differences in similarity of faces encoded during refresh, which could otherwise explain differences in final recognition performance following retrieval and re-study. This experimental paradigm thus further strengthens the conclusion that any differences in memory following retrieval versus re-study can be attributed to differential engagement of active retrieval processes.

However, there is one important caveat with the results from Experiment 4a. Despite the consistency with published research, the findings from Experiment 4a could be explained by participants becoming fatigued during the second block, or some other order effect (such as interference) that influenced block two performance in a general way. That is, all measures during final recognition were lower in block one than block two (see Figures 4.5 and 4.6). Although this pattern was predicted for final recognition accuracy and bias measures for repeated trials, differences in final recognition accuracy for the baseline condition were not expected. Therefore, it is possible that reductions in final recognition accuracy and bias measures across blocks were due to participant fatigue, or other order-related effects, from block one to block two. An order effect is problematic for interpreting the present experiment given that the refresh task in block two was always the re-study task. Consequently, a second experiment was conducted to further explore whether the results from Experiment 4a could be explained by an order effect of block. For Experiment 4b, the refresh task in both block one and two was an active retrieval task, meaning that both blocks were identical to one another with the exception of the face stimuli shown, and the order of block. Given that face stimuli were counter-balanced between blocks, any differences in final recognition performance in this design can only be attributed to order effects as the experiment progressed.

4.2. Experiment 4b

Experiment 4b had a similar design to Experiment 4a, where participants completed two blocks of learning, filler, refresh, filler and final recognition tasks. In both blocks, participants learnt 30 targets before completing a retrieval task during refresh for 20 of the 30 face targets. For each trial, participants were presented with one target face and four distractor faces from the same face set as the target. Participants were asked to select the face they recognised as one shown during learning. The same retrieval task was completed during both block one and block two refresh. For both blocks, the final recognition task required participants to complete the same recognition task done in refresh, but for all 30 face targets (20 repeated trials, 10 baseline trials). Trial timings, stimuli, and instructions were identical to Experiment 4a, with the only exception that both blocks involved active retrieval instructions.

From this paradigm, and to provide direct comparisons to results from Experiment 4a, analysis of Experiment 4b data focused on target selection rates across refresh and final recognition, as well as face memory updating measures during the final recognition task. Refresh target selection measured the extent to which participants correctly recognised target faces during the initial refresh task. Similarly, final recognition target selection measured the extent to which participants correctly recognised the target face, separately for repeated trials and baseline trials. Furthermore, for refresh incorrect trials only, final recognition bias measured the extent to which participants repeated the same recognition error during final recognition as opposed to switching to a different recognition error (using the same formula as in Experiment 4a). All measures were calculated separately for block one and two. It was hypothesised that face recognition target selection and bias measures during final recognition should be similar across block, given that both blocks involved a retrieval task during refresh. Alternatively, if the performance differences across blocks in Experiment 4a were due to order effects such as fatigue, then final recognition accuracy and bias should be higher in block one compared to block two.

4.2.1. Method

Participants

Seventy-two undergraduate participants, 62 female (M_{age} = 19.29, SD_{age} = 1.52) and with an age range of 18-28, were recruited from the University of Kent. This sample size was recruited to match the sample size of Experiment 4a. With this sample size, there was 99% power to detect a medium effect size (0.5) with a two-tailed alpha of .05 and 71%

power to detect a small effect size (0.3) with a two-tailed alpha of .05. All participants were naïve to the experiment aims, and were not recruited for Experiment 4a nor any other experiment from this thesis. Participants received course credits for participation and were treated in line with ethical approval received from the University of Kent ethics board. Participants were required to have normal/corrected to normal vision and be aged between 18-35 years old. All recruited participants were eligible for analysis.

Stimuli

Face stimuli. The face stimuli used in Experiment 4b were the same 60 face sets and images used in Experiment 4a.

Apparatus, procedure and design

The apparatus and procedure for Experiment 4b was identical to Experiment 4a with the exception of the refresh task completed in block two. In block one (see Figure 4.7), participants completed a learning phase for 30 face targets, before completing a short letter search filler task. After this, an active retrieval task was done where, for each trial, participants were presented with one target face and four distractor faces from the same face set as the target. Participants selected the face they recognised as one shown during learning. Following a second letter search filler task, the same retrieval task was completed during final recognition for the 20 repeated trials (shown during refresh) and 10 baseline trials (not shown during refresh). The exact same task procedure was completed in block two, except that face stimuli were taken from the alternative face set list that was not shown during block one. Therefore, for Experiment 4b, the design was a within-subject design with manipulated factors being the order of blocks (one and two) and the types of trials during final recognition (repeated, baseline). As in Experiment 4a, face stimuli were counter-balanced according to whether face sets were presented during refresh and final recognition (20 repeated trials) or only presented during final recognition (10 baseline trials). Additionally, face set lists were counter-balanced to be presented in block one or two.

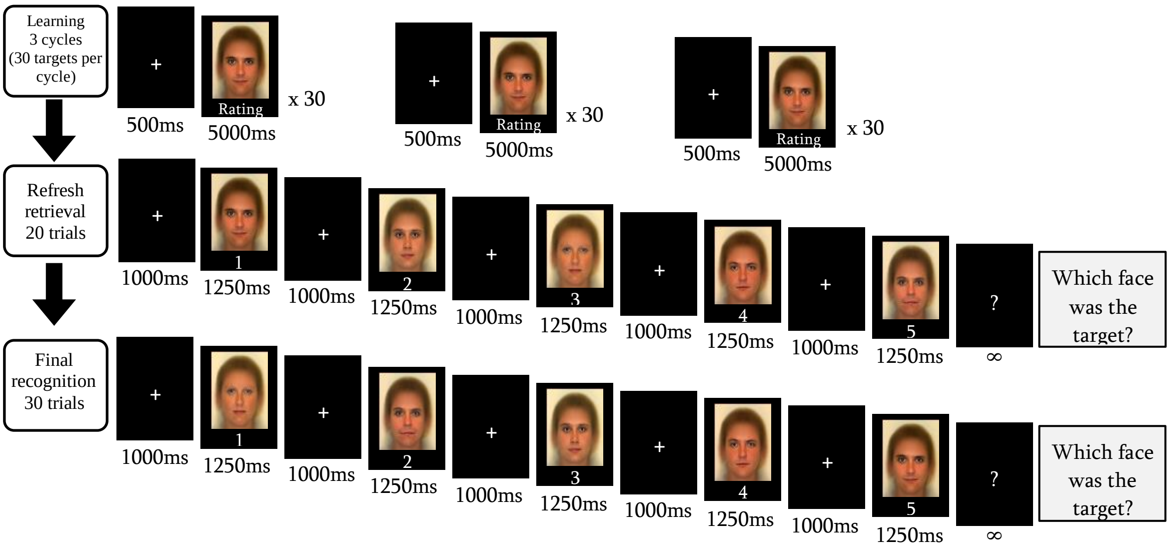

Figure 4.7. Experiment 4b block one and two procedure. During learning, participants encoded 30 targets, with all targets presented in three cycles. Next, participants completed 20 ‘active’ face recognition trials where participants were asked to select a target face shown amongst four distractor faces. The same face recognition task was completed during final recognition, but for all 30 face targets (20 repeated trials, 10 baseline trials). The only difference between blocks was the face set list.

Data analysis

Similar to Experiment 4a, the same measures of final recognition target selection and bias rates were calculated, separately for blocks one and two. Two exceptions were made, however, for Experiment 4b. First, as recognition responses between refresh tasks in blocks one and two could now diverge (as no matching procedure was implemented for active retrieval responses during block one and block two refresh), refresh target selection was statistically compared across blocks. Second, for all reported statistical analysis, both Bayesian and frequentist statistics were calculated. Bayesian statistics were added to Experiment 4b analysis as this experiment tested two competing hypotheses for each final recognition and bias contrast. One strength of Bayesian analysis is that this approach shows the ratio of evidence in favour of one hypothesis over another from the observed data. For each t-test performed in the following analysis, Bayes factors were calculated with informed priors, with the prior centered at the observed effect sizes seen for the equivalent contrast shown in Experiment 4a. For the Bayesian t-test of refresh accuracy between block one and two, a prior effect size was calculated based on the two effect sizes shown for the difference between block on final recognition accuracy measures in Experiment 4a.

4.2.2. Results

Target selections refresh during refresh and final recognition

As in Experiment 4a, data from Experiment 4b was initially analysed for differences in target recognition rates across block, separately during refresh and final recognition (for repeated and baseline trials). As mentioned, analysis of Experiment 4b incorporated Bayesian statistical analysis for each equivalent frequentist analysis, with Bayes Factors larger than 3 supporting the alternative and Bayes Factors less than 0.33 supporting the null hypotheses, for all analyses. All statistical results for each ANOVA and t-test contrast can be found in Table 4.1.

All measures of target selection rates were significantly greater than chance performance (0.2) at *p* <.001 (see Figure 4.8). To analyse the influence of trial condition (repeat, baseline) and block order (block one, block two) for final recognition target selection, a 2 x 2 ANOVA was conducted (see Table 4.1). This analysis showed no main effect of condition, but a main effect of block. Furthermore, no condition by block interaction was observed. Planned contrasts were also run to compare target selection rates for repeated trials between block one and two. As seen in Figure 4.8, target selection rates was significantly greater during block one compared to block two for both repeated and baseline trials, suggesting that differences in final recognition accuracy across blocks were driven by order effects.

In addition, proportion target selections was also compared between block one and two from the refresh phase. This analysis showed that proportion target selections was larger in block one compared to block two refresh (Figure 4.8). Thus, all contrasts of recognition accuracy across block one and two showed that target selection was different across blocks, with both frequentist and Bayesian statistics. Considered together, these results support the hypothesis that order effects account for the observed drop in recognition accuracy across block.

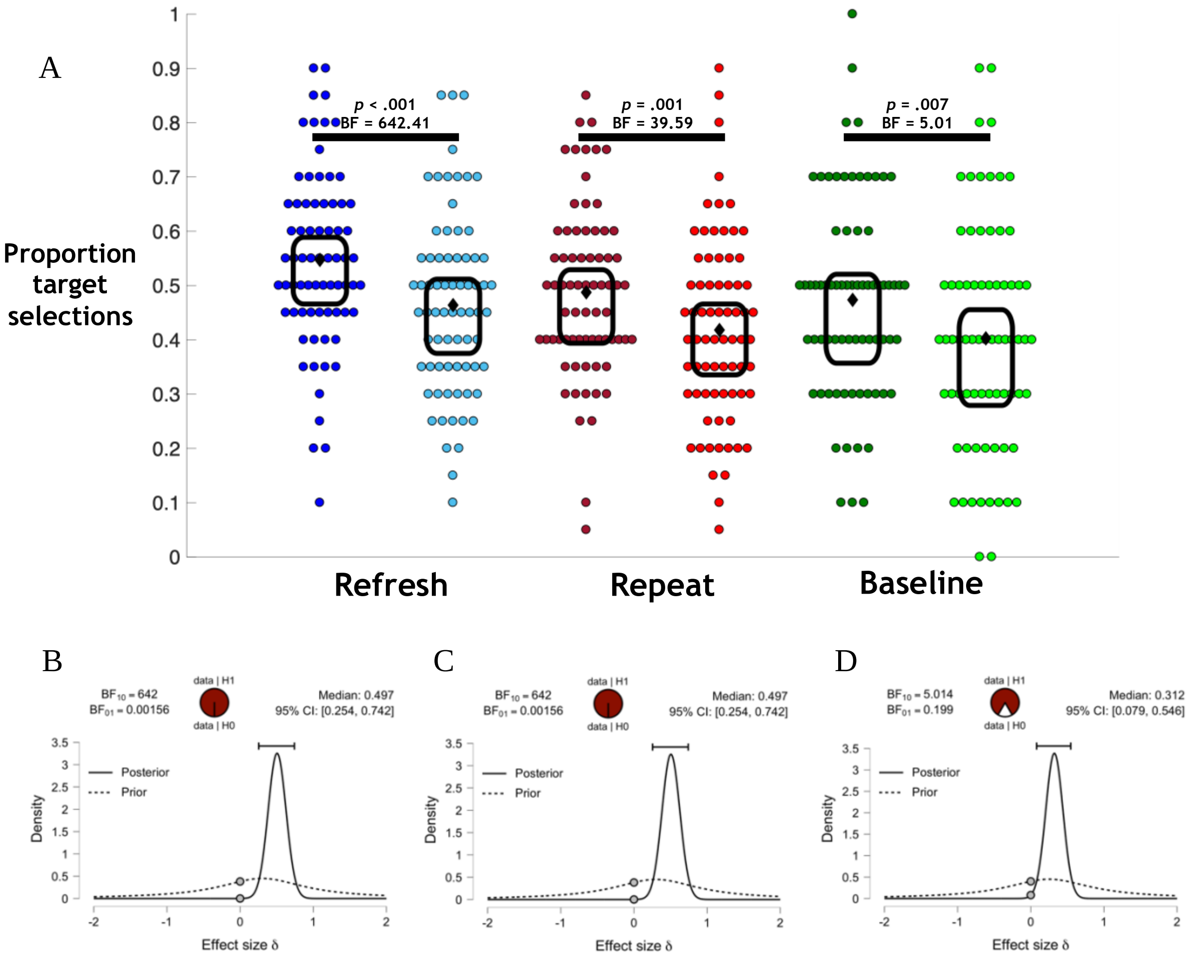

Figure 4.8. Statistics for the analysis target selections during Experiment 4b. Panel A. Proportion target selections during refresh and final recognition for block one (darker colours) and two (lighter colours). Plots of prior and posterior distributions for Bayes Factors of each contrast of target selection rates during refresh (panel B), repeated (panel C) and baseline conditions (panel D).

Table 4.1. Frequentist and Bayesian statistics for each ANOVA and t-test contrast performed for Experiment 4b. Bayes Factors (BF) represent evidence in favour of an alternative hypotheses (values greater than 1), evidence in favour of a null hypothesis (values less than 1) or evidence that is inconclusive (values close to 1).

ANOVA effects	df	F	p	η^2p	BF_{10}	Prior
Trial type	1,71	0.002	.97	<.001	0.18	0.20
Block	1,71	18.93	<.001	0.21	880.72	0.20
Trial type * Block	1,71	0.87	.35	0.01	0.17	0.20

Pairwise differences (block 1 vs. 2)	df	t	p	g	BF_{10}	Prior
Refresh	71	4.38	<.001	0.52	642.41	0.30
Final test, repeated	71	3.50	<.001	0.41	39.59	0.36
Final test, baseline	71	2.76	.007	0.33	5.01	0.25
Final bias difference measure	71	1.06	.29	0.13	0.21	0.33

Final recognition bias comparison between block one and two

Next, the final recognition bias difference measure was analysed for block one and two[2]. A more positive bias difference score indicates that participants were more likely to repeat recognition responses, rather than switch between distractor recognition responses, during final recognition. The final recognition bias difference score, as demonstrated in Figure 4.9, was significantly higher than 0 in blocks one ($t(71) = 6.35$, $p <.001$, $g = 0.75$) and two ($t(71) = 6.14$, $p < .001$, $g = 0.72$), suggesting that both refresh retrieval tasks resulted in encoding of face distractors during refresh. However, as noted in Table 4.1, no

---

[2] All 72 participants were included in the bias analysis. In an alternative analysis (not presented), the same analyses were repeated after removing 7 participants who had less than 0.2 accuracy during refresh retrieval in both blocks (i.e. those who show poor memory for targets). However, the conclusions from both analyses are the same, therefore the whole sample is included to avoid introducing biases in the sample.

significant difference was found in final recognition bias difference measures between block one and two. Furthermore, the Bayesian statistics for this contrast confirmed that evidence favoured the null over the alternative hypothesis. These results suggests that both retrieval tasks during refresh induced a similar level of face memory updating during final recognition.

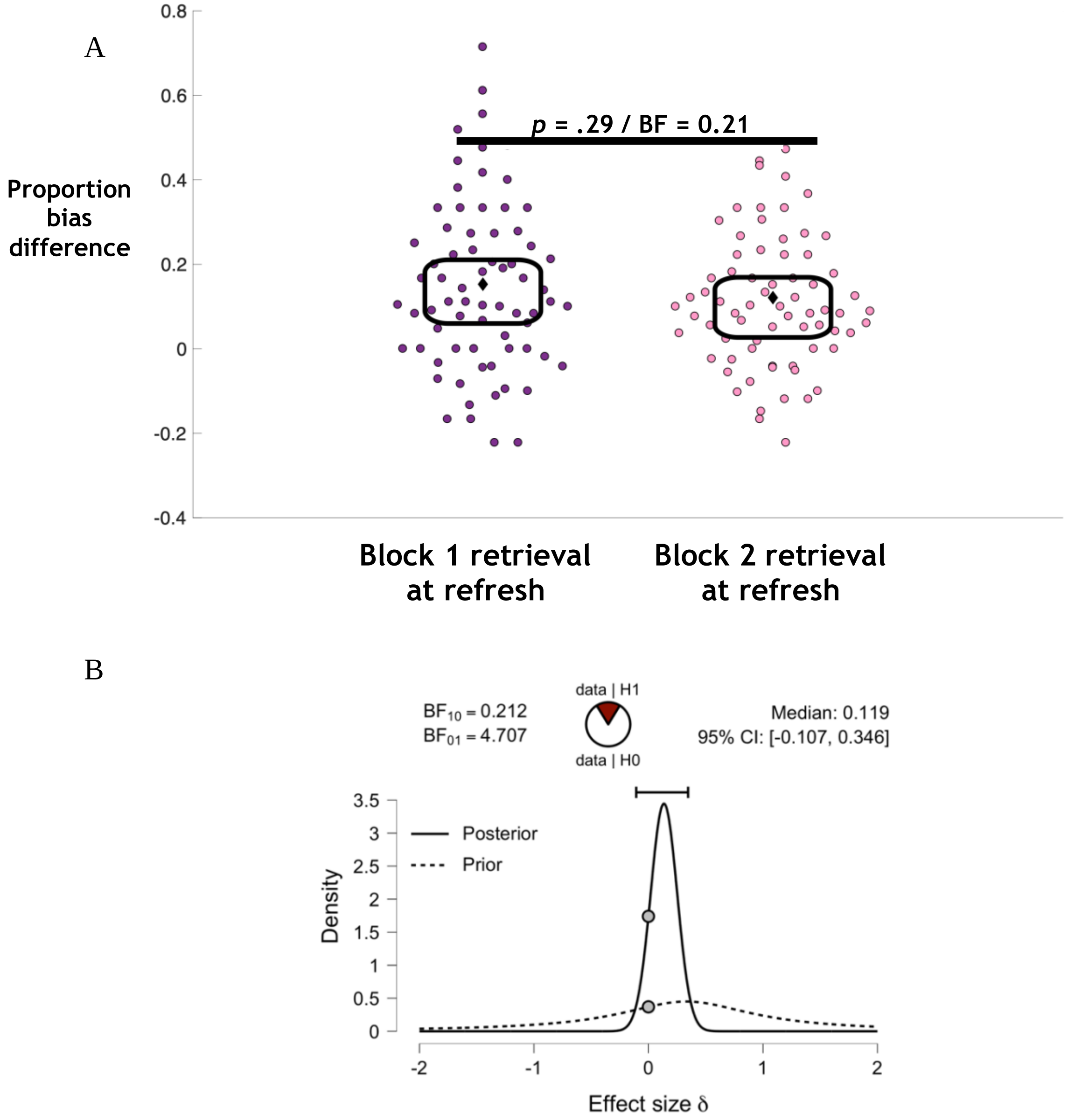

Figure 4.9. Statistics for the analysis of bias difference during Experiment 4b. Panel A. Bias difference measures following retrieval during block one and two refresh. Panel B. Plots of prior and posterior distributions for Bayes Factors of each contrast of final recognition bias difference measures.

4.2.3. Discussion

The aim of the second experiment in this chapter was to establish whether differences in final recognition accuracy and bias in Experiment 4a were due to the manipulation of refresh task, or could be explained by an order effect of block. For Experiment 4b, participants completed two blocks of learning, refresh and final recognition task. For both blocks, participants completed an active retrieval task during refresh, where they had to recognise which of five faces was a target shown during the preceding learning phase. The same task was completed during final recognition, meaning the only systematic difference between blocks was the order in which they appeared (although stimuli were different across blocks, this was counterbalanced across participants so cannot explain group level differences). Using this paradigm, it was tested whether face recognition performance was susceptible to order effects, which would be expected if there was a drop in accuracy/bias from block one to two.

The results suggest a strong order effect was present for face recognition accuracy. That is, participants had significantly higher recognition accuracy during refresh and final recognition, both for repeated and baseline trials, for block one compared to block two. Such results support the argument that recognition accuracy, in the current experimental paradigm, was susceptible to order effects, given that all three measures of recognition accuracy demonstrated a similar drop in accuracy rates across block. In contrast, final recognition bias did not reduce from block one to block two, which was supported by the Bayes factor result indicating evidence in support of the null hypothesis of no difference between blocks. Thus, the bias measure did not seem to be affected by order in the same way as accuracy. This finding supports the hypothesis that updating of face memories is more likely following active retrieval attempts, rather than passively studying face images such as in Experiment 4a (Roediger & Marsh, 2005; Bridge & Voss, 2014).

Taken together, the results from Experiment 4b showed support for order effects on accuracy, but support against order effects on bias. In order to explain why final recognition accuracy and bias were differentially affected by order, we must consider the way in which these measures were calculated. Whereas final recognition accuracy was calculated from all 20 repeated trials, final recognition bias was calculated only from trials where an error was made during the previous refresh trial. Studies from the retrieval practice effect literature have shown how retrieval practice benefits can vary according to whether data is conditionalised on initial retrieval/re-study accuracy (Rowland & Delosh, 2014). In the current study, recognition bias was calculated for trials involving refresh errors only, as this was considered the purest measure of retrieval bias and memory updating that avoided potential accuracy confounds between conditions. These two measures where therefore statistically decoupled, which allowed testing of order effects separately for each measure. Reduced accuracy in block two compared to block one could have been caused by less efficient encoding processes during the initial learning phase, or less efficient retrieval processes during the final test phase as a result of factors such as fatigue or cumulative interference between similar memories. In contrast, bias was not affected, suggesting that encoding of errors during refresh was equal across both blocks.

Whilst the data from Experiments 4a and 4b provides support that retrieval is a stronger inducer of face memory updating than re-study, one potential confound with the reported paradigm is that retrieval and re-study tasks during refresh were not matched on participants’ freedom of choice during refresh selections. In other words, retrieval during refresh required participants to choose a face they recognised to be the target face, with this choice not constrained to one particular face image. In contrast, re-study during refresh constrained participants to study a face that was indicated by the matching procedure. This choice confound is problematic as participants in the re-study condition may have recognised faces within the trial that were then not cued to be encoded for subsequent retrieval tests. Furthermore, previous research on the “self-choice” effect suggests that free choice has an enhancing effect on memory encoding. The self-choice effect refers to the finding that participants typically have better memory for stimuli they have chosen to encode during a learning phase, compared to stimuli that were chosen by the experimenter for participants to learn (Watanabe, 2001; Watanabe & Soraci, 2004). Furthermore, Murty, DuBrow and Davachi (2015) demonstrated the self-choice effect was present even when choices during learning were irrelevant to content to-be-remembered (as items to-be-remembered were masked by symbols that, once selected, revealed the items). These self-choice effects were accompanied by enhanced hippocampal-striatal interactions during encoding, suggesting that self-choice of stimuli in a memory encoding paradigm boosts subsequent retrieval performance by engaging brain networks supporting learning.

According to retrieval-induced updating theories (e.g. Bridge & Paller, 2012), encoding mechanisms are engaged during retrieval, meaning that during subsequent retrieval, participants remember information that was selected during the previous retrieval attempt, rather than information that was originally encoded. However, encoding mechanisms may be engaged through the act of choosing, rather than through the process of retrieval. That is, any enhancement of face memory updating following retrieval versus re-study at refresh may not be specifically due to retrieval attempts, but instead may be due to enhanced learning as a result of choosing a response. In the current study, that would have involved choosing one of the five images (during retrieval) versus being constrained to focus on encoding one of the five faces (during re-study). However, the self-choice effect has only been examined in the context of initial memory encoding of novel stimuli (Murty et al., 2015; Watanabe, 2001; Watanabe & Soraci, 2004), rather than as a post-learning mechanism for memory updating. Therefore, it is not clear from prior literature whether self-choice would contribute to updating of established memories, as investigated in this thesis.

One relevant study from Karpicke and Zaromb (2010), however, compared the effect of a cued recall task (requiring retrieval) versus a ‘generation’ task and re-study task on subsequent retrieval accuracy. Participants initially encoded several word (e.g. love, diet), and participants then completed a cued recall task, a generation task or a read task. For cued recall and generation tasks, participants were presented with word pair fragments (e.g. heart-l_v_, eat-di__). In the cued recall task, participants attempt to recall the word shown during initial encoding, whereas in the generation task, participants were asked to generate a word that that would fit the word fragment (thus involving choosing a response). Thus, these two tasks were differentiated by ‘retrieval mode’ (i.e., intentional retrieval of memories); whereas the cued recall task required intentional retrieval to complete the task, the generation task did not require retrieval, thus any retrieval processes used in this task were incidental. The read task simply required participants to read in-tact word pairs (e.g. heart-love, eat-diet). In a subsequent free recall or recognition task, it was shown that retrieval accuracy was higher following the cued recall task, compared to generation and read tasks. This study thus suggested that active retrieval attempts (retrieval mode) were particularly important for modifying episodic memories as measured at subsequent tests, even when contrasted against a task involving self-choice of a response.

To investigate whether active recognition attempts also enhance updating of face memories compared to self-choice, an adapted version of Experiment 4a was developed, where participants completed a ‘selection’ task during block one refresh, and a passive re-study task in block two refresh. The selection refresh task required participants to choose a face they thought was the most distinctive image. Critically, this task maintains participants’ ability to choose any of the five faces shown, whilst removing the need to use retrieval processes to guide their decision. The selection refresh task should then serve as a critical control condition to determine whether previous memory updating effects from Experiments 4a and 4b are retrieval-induced, or are instead accounted for by self-choice effects, irrespective of whether this choice was based on retrieval or not.

4.3. Experiment 4c

In order to test whether the results from Experiment 4a and 4b could be explained by a self-choice effect, a third Experiment (4c) was conducted that incorporated a ‘selection’ task during refresh. As in previous experiments, participants completed two blocks of learning, refresh and final recognition tasks. Thirty face targets were encoded during the learning phase, before participants completed either a ‘selection’ task (block one) or a re-study task (block two) during refresh. Refresh trials consisted of one target face shown amongst four distractor faces from the corresponding face set to the target, with 20 trials completed in total. For the selection task, participants selected the face they thought was the most distinctive of the five faces, with no mention that they should try to recognise which face was shown during learning. For block two refresh, participants were told which face from the five images was to be learnt for the subsequent memory test. Following both refresh conditions, participants completed a final recognition task, with each trial requiring participants to select the face they recognised from the original learning phase, for all face targets (20 repeated, 10 baseline trials).

In order to compare the results of Experiment 4c to previous experimental results, the same measures from the final recognition task were extracted. Final recognition target selection rate measured the extent to which participants successfully recognised target faces during final recognition, separate for repeated and baseline trials. For repeated trials, final recognition bias measured the extent to which participants would repeat the same recognition error from refresh task to final recognition. As in the two previous experiments, these measures were calculated separately for blocks one and two, to assess the impact of selection (without retrieval) versus re-study on face memory updating.

For the third experiment, two opposing predictions were formulated. If retrieval the only critical factor for face memory updating, then there should be no differences in final recognition bias following select and re-study refresh. In contrast, should the results from Experiments 4a and 4b be explained by a self-choice effect, then final recognition bias should be larger following select versus re-study refresh tasks, consistent with a self-choice enhancing encoding during refresh.

4.3.1. Method

Participants

Seventy-two undergraduate participants ranging in age from 18-28, of which 59 were female (M_{age} = 19.86, SD_{age} = 4.46) were recruited from the University of Kent. This sample size was recruited to match the sample size of Experiment 4a. With this sample size, there was 99% power to detect a medium effect size (0.5) with an alpha of .05 and 71% power to detect a small effect size (0.3) with a two-tailed alpha of .05. All participants were naïve to the experiment aims and were not recruited for any other experiments in this thesis. Participants received course credits and were treated in line with ethical approval received from the University of Kent ethics board. Participants had normal/corrected to normal vision and were aged between 18-35. All recruited participants were eligible for analysis.

Stimuli

Face stimuli. The face stimuli used in Experiment 4c were the same 60 face sets and images used in Experiment 4a and b.

Apparatus, procedure and design

The apparatus and procedure for Experiment 4c was identical to Experiments 4a and 4b, with the exception of the refresh tasks completed in blocks one and two. In block one (see Figure 4.10), participants learnt 30 face targets, before completing a short letter search filler task. After this, participants completed a select refresh task. For each of the 20 trials, participants were shown five faces, one target face and four distractor faces from the same face set as the target. After seeing all five images, participants selected the face they thought was most distinctive, and were told to memorise this face for the subsequent memory test. The selection task was similar to the retrieval task in that participants were instructed to focus on all five faces before making a retrospective judgement on which face they thought was the most distinctive. Following a second short filler task, participants completed 30 final recognition trials, including 20 repeated trials (shown during refresh) and 10 baseline trials (not shown during refresh). For each trial, participants were shown a target face amongst four distractor faces, and were told to select which face they recognised from the original learning task. The same procedure was completed in block two (see Figure 4.11), except that during block two refresh, participants completed a passive re-study task. This task was identical to that completed in Experiment 4a, such that the face participants were asked to encode (cued by a white fixation cross) was matched to the responses made by participants in the preceding select refresh task, with the difference that those responses were now based on selecting a distinctive face rather than selecting a recognised face. The only other difference between blocks one and two was the face set list presented. As in Experiments 4a and 4b, face stimuli were counter-balanced according to whether face sets were presented during refresh and final recognition (20 repeated trials) or only presented during final recognition (10 baseline trials). Additionally, face set lists were counter-balanced to be presented in block one or two. With Experiment 4c, a within-subject design was adopted with manipulated factors of refresh task (selection, passive re-study) and trial type (repeated, baseline).

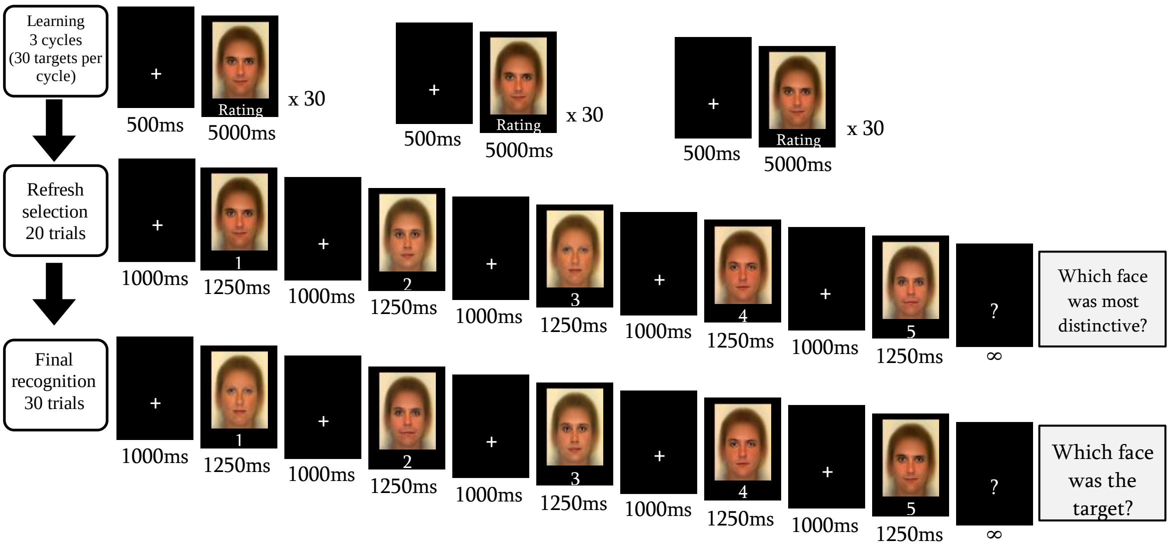

Figure 4.10. Block one procedure in Experiment 4c. During learning, participants encoded 30 targets. Next, participants completed 20 select refresh trials where participants selected which face from five was the most distinctive. During final recognition, participants were asked to select the target face shown amongst four distractor faces for trials shown during refresh (repeated condition) and trials not shown during refresh (baseline condition).

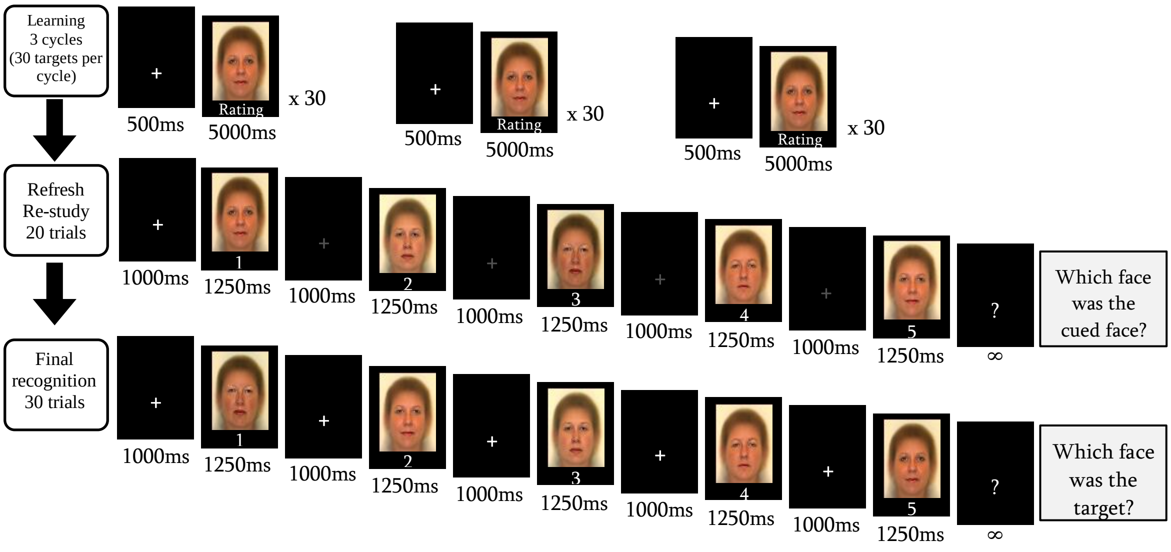

Figure 4.11. Block two procedure in Experiment 4c. During learning, participants encoded 30 targets, with all targets presented in three cycles. Next, participants completed 20 ‘passive’ re-study trials where participants were asked to encode one of five faces. During final recognition, participants were asked to select a target face amongst four distractors for trials shown during refresh (repeated condition) and trials not shown during refresh (baseline condition).

Data analysis

As in Experiments 4a and 4b, the same measures were calculated for Experiment 4c. As a reminder, for the 20 ‘refresh’ trials, refresh target selection was calculated as the proportion of refresh trials where the target face was chosen. Due to the matching procedure adopted in Experiment 4c, refresh target selection rates was identical across blocks one and two. Additionally, final recognition target selection measure for repeated trials calculated the proportion of repeated trials where the target face was selected during final recognition. For repeated trials, the bias difference score was also calculated as the primary measure of face memory updating. Finally, for baseline trials, the proportion of these trials where the target face was selected was also calculated during final recognition.

Similar to the analysis in Experiment 4b, both Bayesian and frequentist statistics were calculated for all analyses. Bayesian statistics were again well suited for the analysis of Experiment 4c as this experiment tested two competing hypotheses for each final recognition target selection and bias contrast. For each t-test performed in the following analysis, Bayes factors were calculated with informed priors, these being the observed effect sizes seen for the equivalent contrast shown in Experiment 4a.

4.3.2. Results

Face recognition target selections during refresh and final recognition

Initial analysis of Experiment 4c data focused on patterns of target selection rates across refresh and final recognition tasks, which were all significantly greater than chance (0.2) at $p < .001$. In addition, as demonstrated in Figure 4.12, proportion target selections during refresh was equivalent between blocks one and two owing to the matching procedure employed between refresh tasks. As expected, the proportion of target selections in Experiment 4c was substantially lower than in previous experiments (where proportion target selection during refresh was around to 50%). These results support the view that participants were not simply choosing faces based on active recognition of target face memories during the selection refresh task.

Next, final recognition target selection was analysed according to trial type (repeat, baseline) and block order (block one, block two). A 2 x 2 ANOVA was conducted on proportion final recognition target selection measures, both with frequentist and Bayesian statistical analyses. As a reminder, Bayes Factors larger than 1 indicate more support for the alternative hypothesis whilst values less than 1 indicate more support for the null hypothesis. These analyses, shown in Table 4.2, showed no main effect of trial type, but a main effect of block. Furthermore, no interaction effect between trial and block was seen. Planned contrasts continued the pattern of findings from Experiments 4a and 4b, whereby final recognition target selection rate was significantly larger in block one versus block two, for both repeated and baseline trials. The Bayes factors for these results, found in Table 4.2, were not influenced by the choice of prior (see Figure 4.12), and provide strong support in favour of the hypothesis that order effects can explain the drop in final recognition accuracy from block one to two.

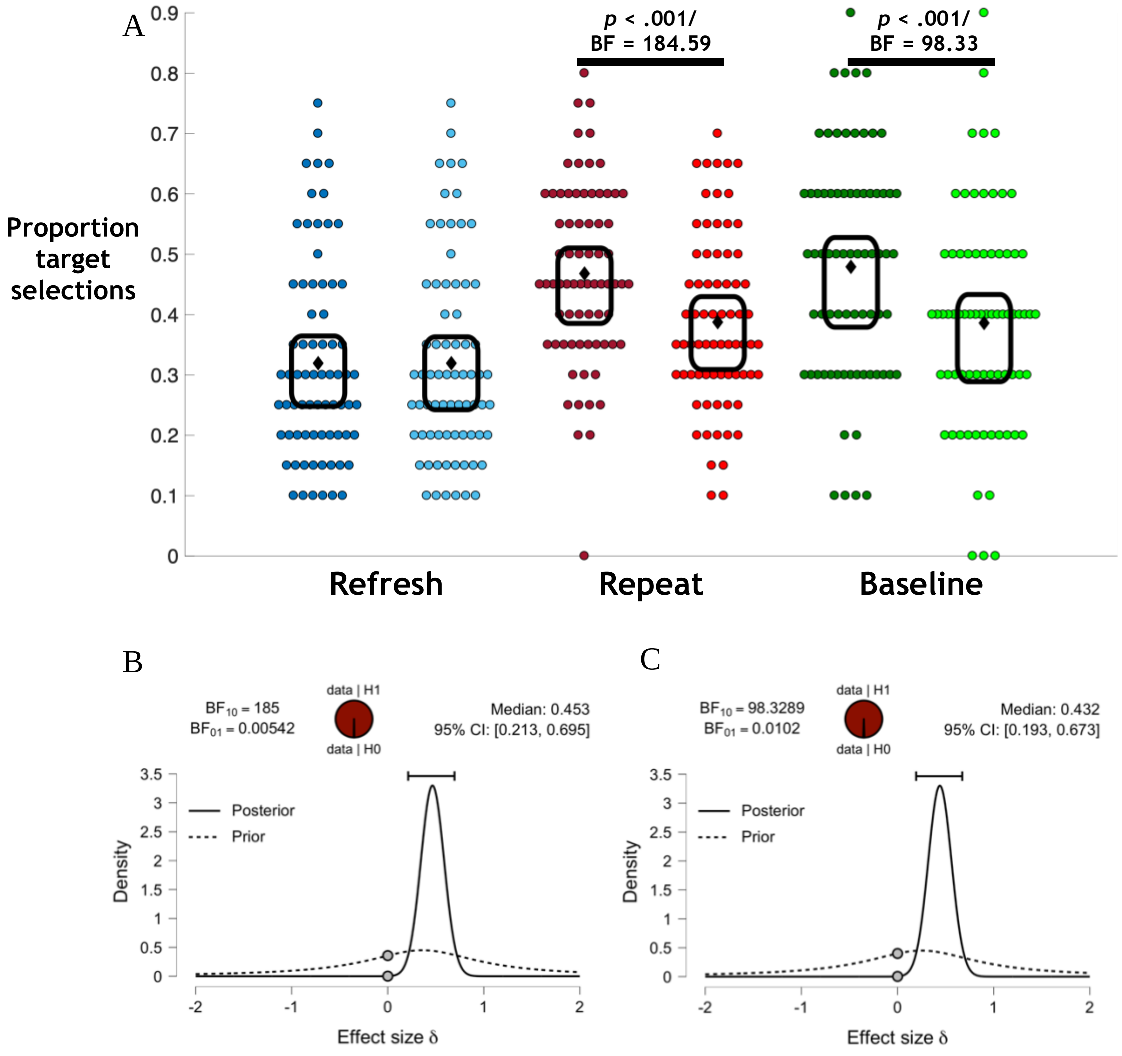

Figure 4.12. Statistics for the analysis target selections during Experiment 4c. Panel A. Proportion target selections during refresh and final recognition for block one (darker colours) and two (lighter colours). Plots of prior and posterior distributions for Bayes Factors of each contrast of target selection rates during test 2 repeated (panel B) and baseline conditions (panel C).

Table 4.2. Frequentist and Bayesian statistics for each ANOVA and t-test contrast performed for Experiment 4c. Bayes Factors (BF) represent evidence in favour of an alternative hypotheses (values greater than 1), evidence in favour of a null hypothesis (values less than 1) or evidence that is inconclusive (values close to 1).

ANOVA effects	df	*F*	*p*	η^2p	BF$_{10}$	Prior
Trial type	1,71	0.21	.65	0.003	0.13	0.20
Block	1,71	24.23	<.001	0.25	898706.91	0.20
Trial type * Block	1,71	0.16	.69	0.002	0.20	0.20

Pairwise comparisons (block 1 vs. 2)	df	*t*	*p*	*g*	BF$_{10}$	Prior
Final test, repeated	71	3.99	<.001	0.55	184.59	0.36
Final test, baseline	71	3.81	<.001	0.50	98.33	0.25
Final test bias difference	71	1.53	.13	0.18	0.40	0.33

Final recognition bias equal following select and re-study refresh tasks

The next analysis focused on the final recognition bias difference score following the select and re-study refresh tasks[3]. As a reminder, the more positive bias difference score indicates that participants were more likely select the same distractor face during final recognition to the one chosen during the select/re-study refresh tasks. The final recognition bias difference score, as demonstrated in Figure 4.13, was significantly higher than 0 following select refresh (*t*(71) = 2.93, *p* = .005, *g* = 0.35), but was not significantly different to 0 following re-study refresh (*t*(71) = 1.30, *p* = .20, *g* = 0.15). These results

---

[3] All 72 participants were included in the bias analysis. No alternative analysis were conducted based on memory performance during refresh as neither refresh tasks measured participants memory for targets, therefore the whole sample is included.

indicate that, whereas participants were more likely to be biased towards previously selected distractors following the select refresh task, participants were equally likely to select any of the four distractor faces during final recognition, following the re-study refresh task.

Focusing on the difference between bias measures between block one and two, paired samples t-tests showed no significant difference (see Figure 4.13). As seen in Table 4.2, the Bayes factor for this contrast provides anecdotal evidence in favour of the null versus alternative hypothesis. Subsequently, given relatively more support for the null hypothesis (although the Bayes Factor was only anecdotal), these findings provide tentative support in favour of the prediction that face memory updating is equally likely following two separate refresh tasks that did not incorporate active retrieval processing.

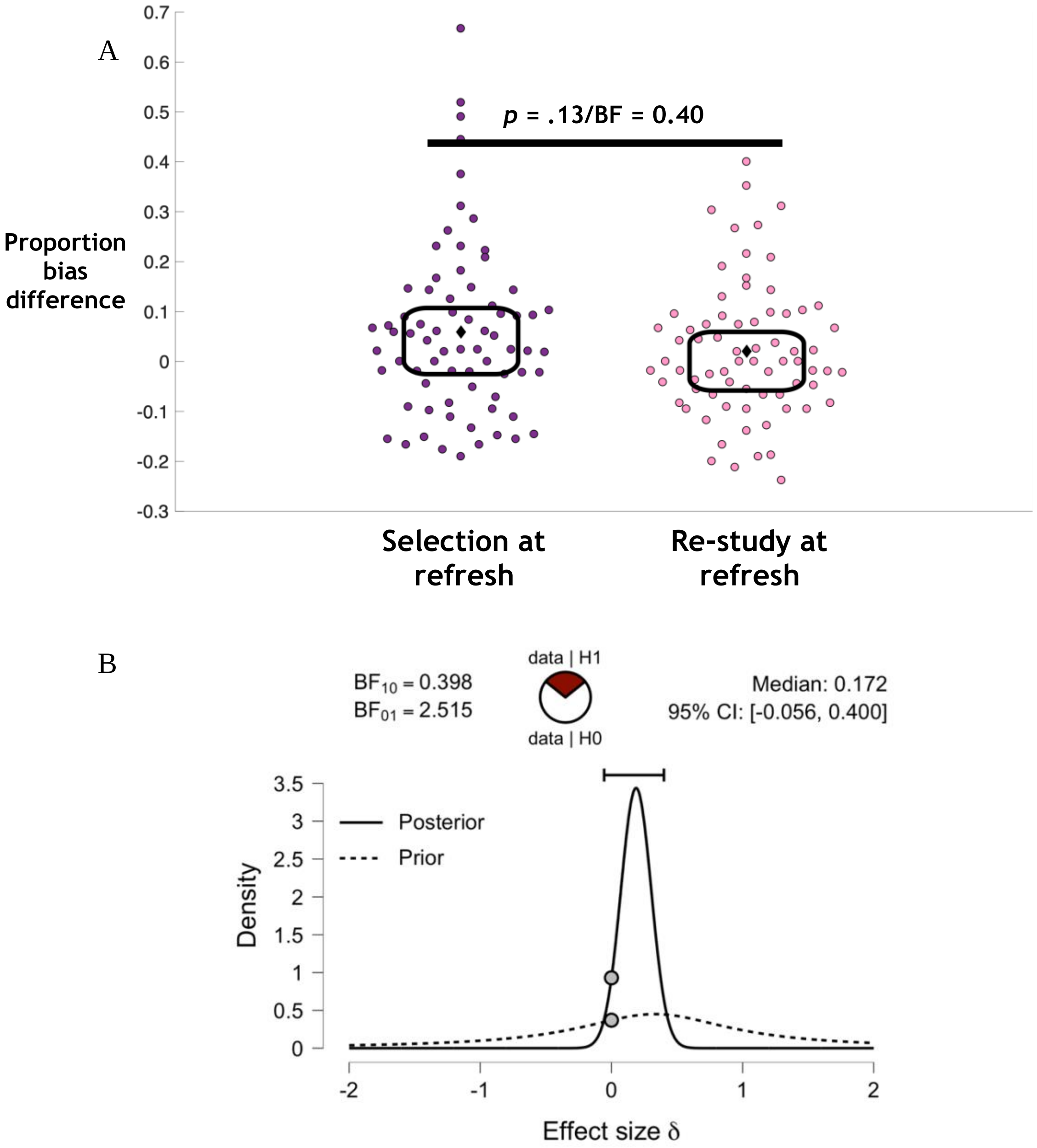

Figure 4.13. Statistics for the analysis of bias difference during Experiment 4b. Panel A. Bias difference measures following retrieval during block one and two refresh. Panel B. Plots of prior and posterior distributions for Bayes Factors of each contrast of final recognition bias difference measures.

4.3.3. Discussion

For the third experiment of the present chapter, it was tested whether enhanced bias after active recognition attempts in previous studies was due to a confound of enhanced self-choice effects for retrieval compared to re-study tasks during refresh. For Experiment 4c, participants completed two blocks of learning, refresh task and final recognition task. For block one refresh, participants completed a ‘selection’ task, whereby the most distinctive face from the five options was chosen. Crucially, the selection task maintained participants’ self-choice during block one refresh, whilst reducing the need for retrieval processes to guide the choices made during this task. During block two refresh, a passive re-study task required participants to encode a face that was chosen for them, based on the type of response made during the matched trial in block one refresh. Following both refresh tasks, final recognition measures of correct target selections and repeated recognition errors provided indicators of face memory updating.

As in previous experiments, final recognition target selection was expected to be larger following select refresh versus passive re-study, for both repeated and baseline trials, owing to an order effect of block on memory accuracy. In contrast, it was anticipated that if the differences in bias across conditions in previous experiments were due to a self-choice confound, final recognition bias should be larger following select refresh compared to passive re-study at refresh. However, if retrieval is a stronger inducer of face memory updating, then final recognition bias should be similar following select and passive refresh. The results from Experiment 4c showed support for both these predictions. Specifically, it was shown that final recognition accuracy was significantly higher following select versus passive re-study at refresh, for both repeated and baseline trials, supporting the claim that overall retrieval accuracy during final recognition was influenced primarily by block order only. In contrast, the final recognition bias difference score was not significantly different following select versus passive re-study. Interpreting Figure 4.13, the average final recognition bias difference scores following select and re-study tasks appear to be similar to the bias score measure following re-study in Experiment 4a (Figure 4.6), and thus substantially lower than bias difference measures observed following retrieval during refresh. Thus, these interpretations support the argument that retrieval, in comparison to re-study tasks that manipulate self-choice, is a stronger mechanism to induce the updating of face memories. However, since self-choice and retrieval refresh manipulations were not directly compared, this conclusion cannot be claimed with certainty without further analyses. In order to provide a direct comparison of final recognition accuracy and bias following retrieval, re-study and selection tasks during refresh, a follow-up between-experiment analysis was conducted to verify that retrieval during refresh is a stronger mechanism for face memory updating, compared to both the re-study and selection refresh conditions.

4.4. Across experiment analysis

In order to compare the final recognition measures across experiment, mixed factorial ANOVA’s were conducted (separately for final recognition target selection and final recognition bias difference). For both measures, two mixed ANOVA’s and planned independent samples t-tests were conducted. First, a mixed ANOVA compared Experiment 4a versus Experiment 4b to test a retrieval-induced updating argument against an order effect argument. Planned t-tests were next conducted to compare final recognition measures between experiments, separately for blocks one and two in order to control for fatigue/order effects. That is, the effect of retrieval during refresh (Experiment 4a) was compared to retrieval during refresh (Experiment 4b) for block one trials. Additionally, the effect of re-study during refresh (Experiment 4a) was compared to retrieval during refresh (Experiment 4b) for block two trials. If memory updating differences between Experiments 4a and 4b were caused by the refresh manipulation, then final recognition measures should be no different for the block one contrast, but should be significantly different for the block two contrast. If order effects account for memory updating effects, then no differences should be found for both block one and two contrasts.

Second, another mixed ANOVA compared Experiment 4a versus Experiment 4c in order to test a retrieval-induced updating argument against a self-choice effect argument. Planned t-tests following this ANOVA was conducted to compare final recognition measures between experiments, again separately for blocks one and two to control for fatigue/order effects. Therefore, the effect of retrieval during refresh (Experiment 4a) was compared to selection during refresh (Experiment 4c) for block one trials. Additionally, the effect of re-study during refresh (Experiment 4a) was compared to re-study during refresh (Experiment 4c) for block two trials. If active retrieval attempts versus self-choice produced different memory updating effects between Experiments 4a and 4c, regardless of block order, then final recognition measures should be different for the block one contrast, but should not be different for the block two contrast. If self-choice effects account for the observed memory updating findings from both active retrieval and self-choice tasks, then no differences should be found for either block one or two contrasts.

For both across-experiment analyses, both alternative and null hypotheses were formulated based on the opposing arguments specified above. Therefore, Bayesian statistics for each analysis were well-suited to provide evidence in favour of alternative versus null hypotheses. Informed priors for t-test contrasts were again used to calculate Bayes factors, with prior distributions for the alternative hypotheses based on effect sizes observed from the alternative experiment analysis, i.e. use effect sizes from Experiment 4c as a prior for the contrast of Experiment 4a vs. 4b, and vice versa for Experiment 4a vs. 4c (so as to not have priors defined from the same dataset that Bayes Factors are calculated from).

Target selections during final recognition

Final recognition target selection measures were analysed as a function of experiment, trial type and block order. For the analysis of Experiment 4a versus 4b, and Experiment 4a versus 4c, see Table 4.3 for ANOVA and Table 4.4 for t-test results.

Experiment 4a versus Experiment 4b. Firstly, a 2 (Experiment; 4a, 4b) x 2 (block; one, two) x 2 (trial type; repeated, baseline) mixed-measures ANOVA was initially done on final recognition accuracy measures. The only significant effect was a main effect of block. As can be seen in Figure 4.14, final recognition target selections was higher in block one compared to block two, regardless of trial type (repeated or baseline trials) and Experiment (4a or 4b). Therefore, these findings support the conclusion that, when comparing retrieval versus re-study at refresh, differences in final recognition accuracy were driven by order of block.

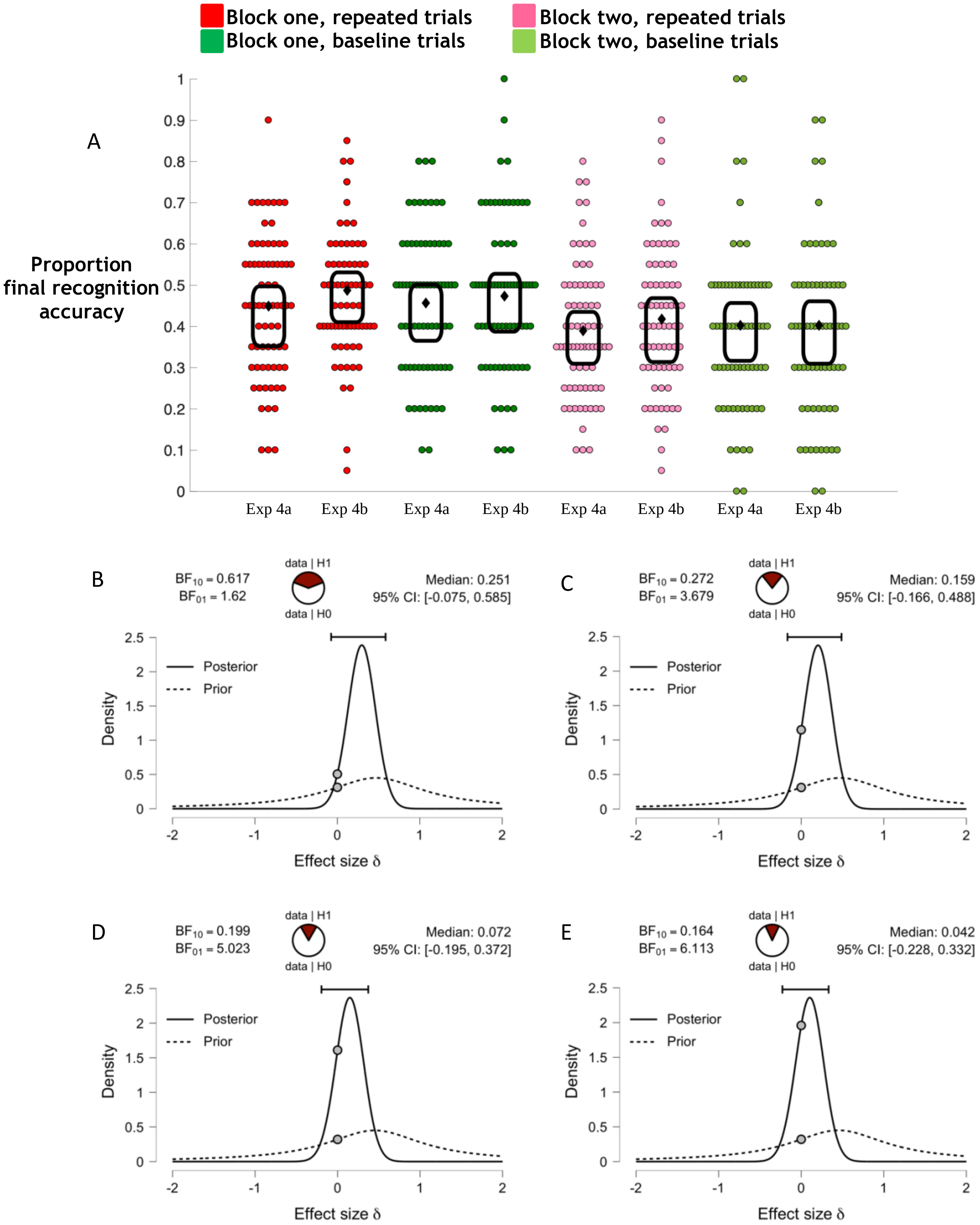

Figure 4.14. Final recognition target selections for Experiment 4a and 4b. Data were analysed for block one (active retrieval blocks) and block two (re-study versus retrieval) for both repeated or baseline trials (panel A). Bayes factors confirmed support for the null hypothesis comparing Experiment 4a versus 4b for block one repeated (panel B), block one baseline (panel C), block two repeated (panel D) and block two baseline trials (panel E).

Table 4.3. Statistical results for mixed-model ANOVA's comparing Experiment 4a versus 4b, and Experiment 4a versus 4c, on final recognition target selection rates (proportion accurate responses). All prior probabilities for each effect were 0.20.

	Experiment 4a vs 4b					Experiment 4a vs 4c			
ANOVA effect	df	*F*	*p*	η^2p	BF	*F*	*p*	η^2p	BF
Trial	1,142	0.03	.86	<.001	0.09	0.49	.49	0.003	0.11
Block	1,142	28.91	<.001	0.17	142306.39	33.88	<.001	0.19	2.560e+7
Experiment	1,142	0.88	.35	0.006	0.25	0.07	.79	<.001	0.17
Trial*Block	1,142	0.01	.93	<.001	0.13	0.03	.87	<.001	0.12
Trial*Experiment	1,142	1.09	.30	0.008	0.25	0.07	.80	<.001	0.14
Block*Experiment	1,142	0.31	.58	0.002	0.15	1.46	.23	0.01	0.30
Trial *Block*Experiment	1,142	0.03	.88	<.001	0.20	0.20	.66	0.001	0.20

Table 4.4. Statistical analysis for paired samples t-tests comparing Experiment 4a versus 4b, and Experiment 4a versus 4c, on final recognition target selection rates. Bayes Factors (BF) represent evidence in favour of an alternative hypotheses (values greater than 1), evidence in favour of a null hypothesis (values less than 1) or evidence that is inconclusive (values close to 0).

		Experiment 4a vs 4b					Experiment 4a vs 4c				
Trial condition	df	*t*	*p*	*g*	BF_{10}	Prior	*t*	*p*	*g*	BF_{10}	Prior
Repeated block 1	142	1.40	.16	0.23	0.62	0.47	0.71	.48	0.12	0.16	0.41
Repeated block 2	142	1.02	.31	0.17	0.27	0.47	0.08	.94	0.01	0.14	0.41
Baseline block 1	142	0.56	.58	0.09	0.20	0.45	0.74	.46	0.12	0.18	0.33
Baseline block 2	142	0.00	1.00	0.00	0.16	0.45	0.52	.60	0.09	0.19	0.33

Experiment 4a versus Experiment 4c. Next, a 2 (Experiment; 4a, 4c) x 2 (block; one, two) x 2 (trial type; repeated, baseline) mixed-measures ANOVA was conducted with final recognition target selection rates analysed. Once again, the only significant effect was a main effect of block, as seen in Figure 4.15 and Tables 4.3 and 4.4. This data highlights how final recognition accuracy was higher in block one compared to block two, regardless of trial type (repeated or baseline trials) and experiment (4a or 4c). Similar to the comparison of Experiment 4a versus 4b, these results again highlight how target selection differences were not influenced by retrieval, select or re-study tasks during refresh. Instead, order effects of block one preceding block two can be seen as the main cause of final recognition target selection differences between block one and two.

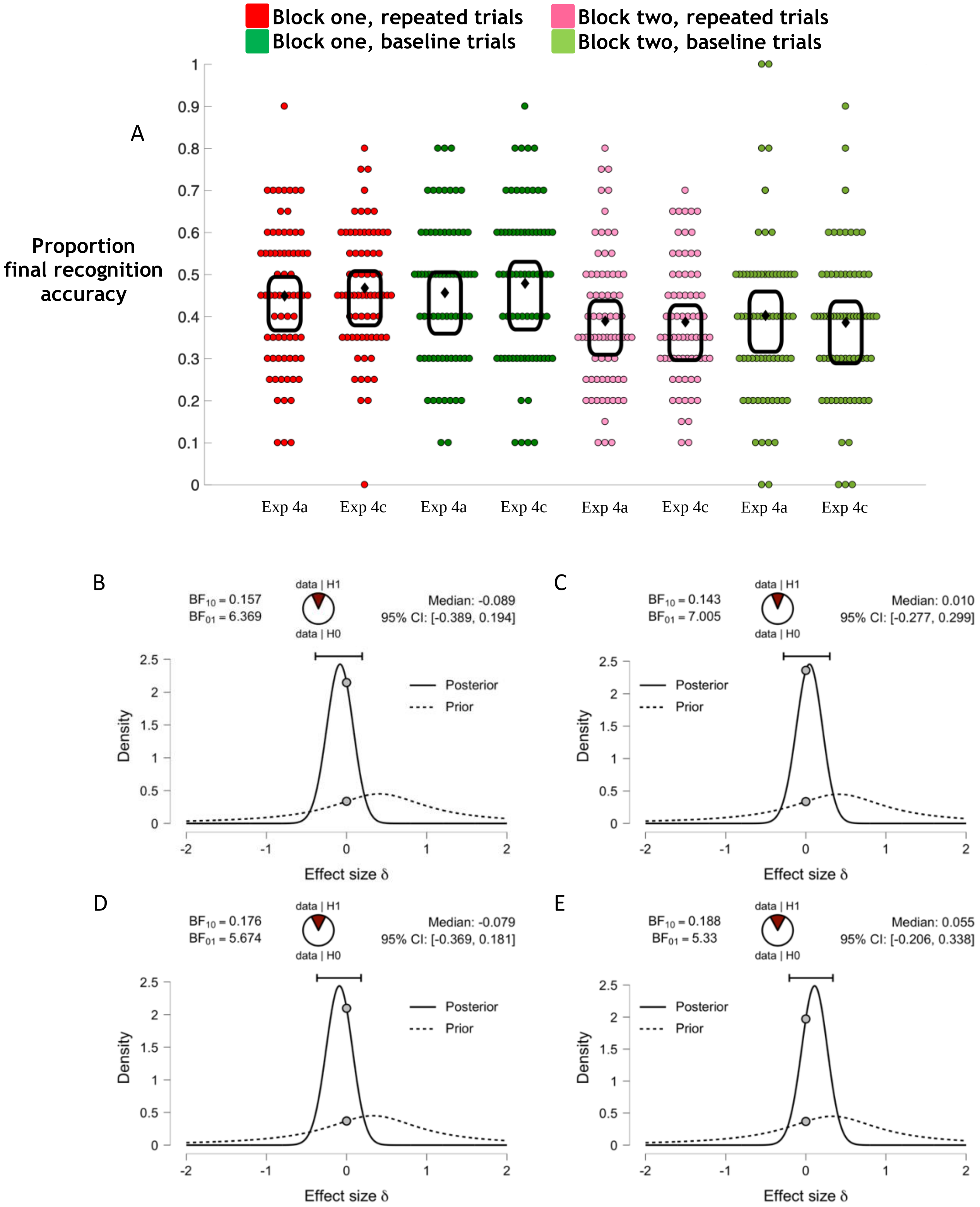

Figure 4.15. Final recognition target selection rates for Experiment 4a and 4c. Data were analysed for block one and block two for both repeated or baseline trials. Bayes factors confirmed support for the null hypothesis in the comparison of Experiment 4a versus 4c for block one repeated (panel B), block one baseline (panel C), block two repeated (panel D) and block two baseline trials (panel E).

Final recognition bias difference

Next, the final recognition bias difference measure was statistically compared as a function of experiment and block presentation. For the analysis of Experiment 4a versus 4b, and Experiment 4a versus 4c, see Table 4.5 for ANOVA and Table 4.6 for t-test results. As a reminder, the contrast of Experiment 4a versus Experiment 4b tested competing predictions that face memory updating is enhanced by retrieval, or is due to a confound of block order. The contrast of Experiment 4a versus Experiment 4c tested the arguments that face memory updating is enhanced by retrieval, or is instead enhanced by self-choice effects irrespective of retrieval processing.

Experiment 4a versus 4b. To assess how final recognition bias difference measure differed between Experiment 4a and 4b, a 2 (Experiment; 4a, 4b) x 2 (block; one, two) mixed-measures ANOVA was conducted[4]. As seen in Table 4.5, a marginal main effect of experiment was found, however a significant main effect of block was seen. Furthermore, the interaction between experiment and block was not significant. Despite this non-significant interaction, planned comparisons were conducted to contrast final recognition bias difference measure between experiments, separately for block one (retrieval vs retrieval at refresh) and block two (re-study versus retrieval at refresh). These contrasts (see Figure 4.16) showed that final recognition bias difference was not significantly different for the block one contrast, as would be expected since both order and refresh task was the same. However, final recognition bias difference was significantly larger for block

---

[4] All 144 participants were included in the bias analysis. In an alternative analysis (not presented), the same analyses were repeated after removing 5 participants from Experiment 4a and 7 participants from Experiment 4b who had less than 0.2 accuracy during refresh retrieval (i.e. those who show poor memory for targets). However, the conclusions from both analyses are the same, therefore the whole sample is included to avoid introducing biases in the sample.

two trials in Experiment 4b (i.e. following active retrieval) versus Experiment 4a (i.e. following re-study). These results support the argument that retrieval is more likely than passive re-study to induce face memory updating.

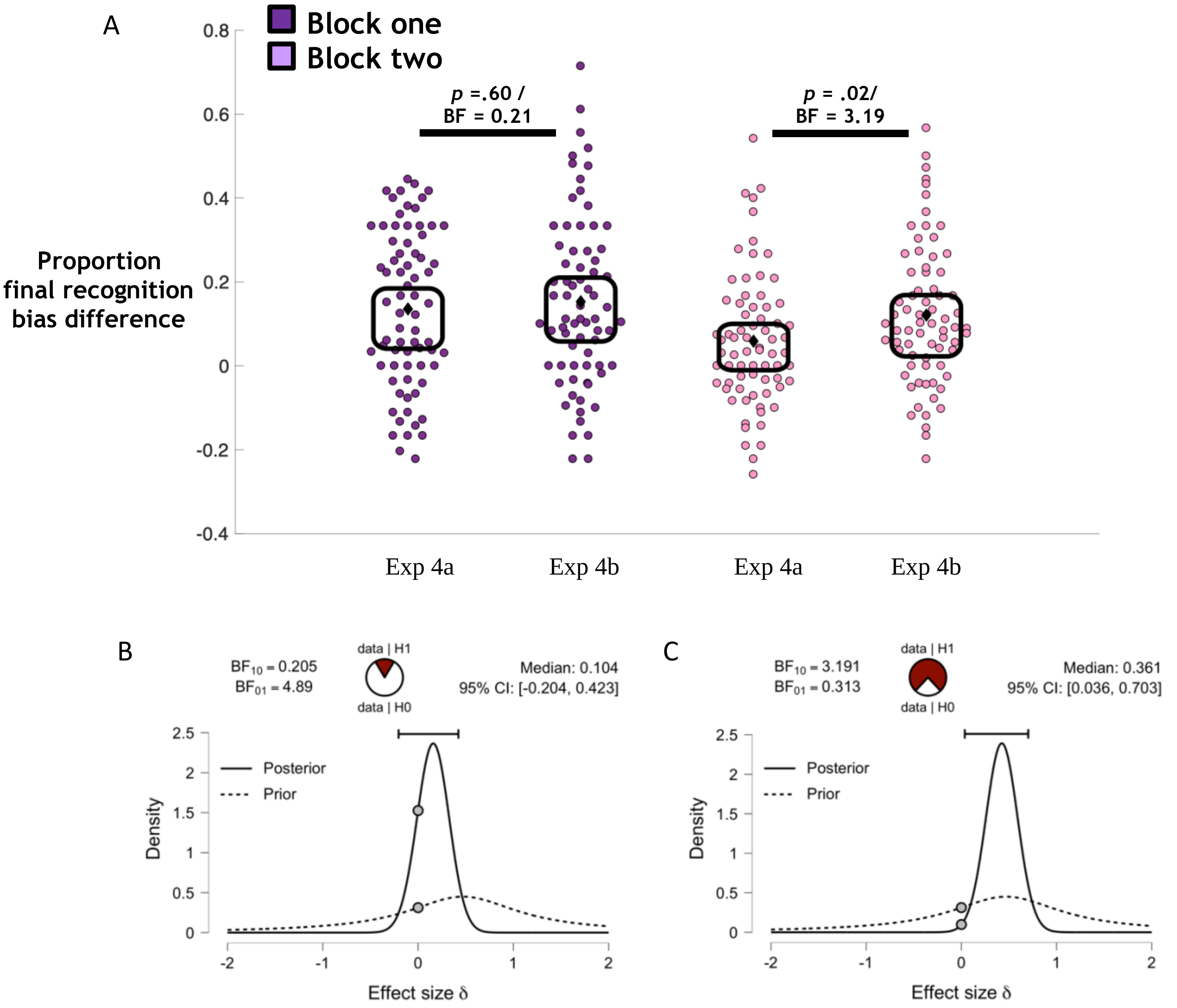

Figure 4.16. Final recognition bias difference from Experiment 4a and b. Final recognition bias was not different following active retrieval at refresh during block one between experiments 4a and 4b (panel B), however bias was significantly increased following retrieval versus re-study during block two between Experiments 4a and 4b (panel C).

Table 4.5. Statistical analysis for mixed-factorial ANOVA’s. ANOVA’s compared Experiment 4a versus 4b, and Experiment 4a versus 4c, on final recognition bias difference. Bayes Factors (BF) represent evidence in favour of an alternative hypotheses (values greater than 1), evidence in favour of a null hypothesis (values less than 1) or evidence that is inconclusive (values close to 1). For Bayes Factors, all prior probabilities were 0.2.

	Experiment 1 vs. 2					Experiment 1 vs. 3			
ANOVA effect	df	*F*	*p*	η^2p	BF	*F*	*p*	η^2p	BF
Block	1,142	7.15	.008	0.05	3.58	9.57	.002	0.06	16.64
Experiment	1,142	3.26	.07	0.02	0.58	8.32	.005	0.06	4.44
Block * Experiment	1,142	1.23	.27	0.009	0.32	1.05	.31	0.007	0.32

Table 4.6. Statistical analysis for paired samples t-tests comparing Experiment 4a versus 4b, and Experiment 4a versus 4c, on final recognition bias difference. Bayes Factors (BF) represent evidence in favour of an alternative hypotheses (values greater than 1), evidence in favour of a null hypothesis (values less than 1) or evidence that is inconclusive (values close to 1).

Pairwise contrast	Experiment 4a vs. 4b						Experiment 4a vs. 4c				
	df	*t*	*p*	*g*	BF	Prior	*t*	*p*	*g*	BF	Prior
Bias difference block 1	142	0.53	.60	0.09	0.21	0.47	2.56	0.01	.43	4.49	0.41
Bias difference block 2	142	2.30	.02	0.38	3.19	0.47	1.57	0.12	.26	0.59	0.41

Experiment 4a versus 4c. Next, the final recognition bias difference measure was analysed between Experiment 4a and 4c in a 2 (Experiment; 4a, 4c) x 2 (block; one, two) mixed-measures ANOVA was conducted[5]. As seen in Table 4.6, both main effects of experiment and block was seen, however there was no significant interaction. Planned comparisons were conducted to compare final recognition bias difference measure between experiment, separately for block one (retrieval vs selection at refresh) and block two (re-study versus re-study at refresh). These contrasts showed that final recognition bias difference was significantly larger in block one for Experiment 4a (i.e. following active retrieval) versus Experiment 4b (i.e. following self-choice without retrieval). On the other hand, final recognition bias difference was not significantly different for block two trials in Experiment 4a versus Experiment 4c (see Figure 4.17), as would be expected, though the Bayes Factor for this contrast provides only anecdotal evidence in favour of the null

[5] All 144 participants were included in the bias analysis. In an alternative analysis (not presented), the same analyses were repeated after removing 5 participants from Experiment 4a who had less than 0.2 accuracy during refresh retrieval (i.e. those who show poor memory for targets). However, the conclusions from both analyses are the same, therefore the whole sample is included to avoid introducing biases in the sample.

hypothesis. These results support the suggestion that self-choice effects alone cannot account for the observed face memory updating effects. Instead, more support was found for the suggestion that retrieval is a critical factor to induce face memory updating.

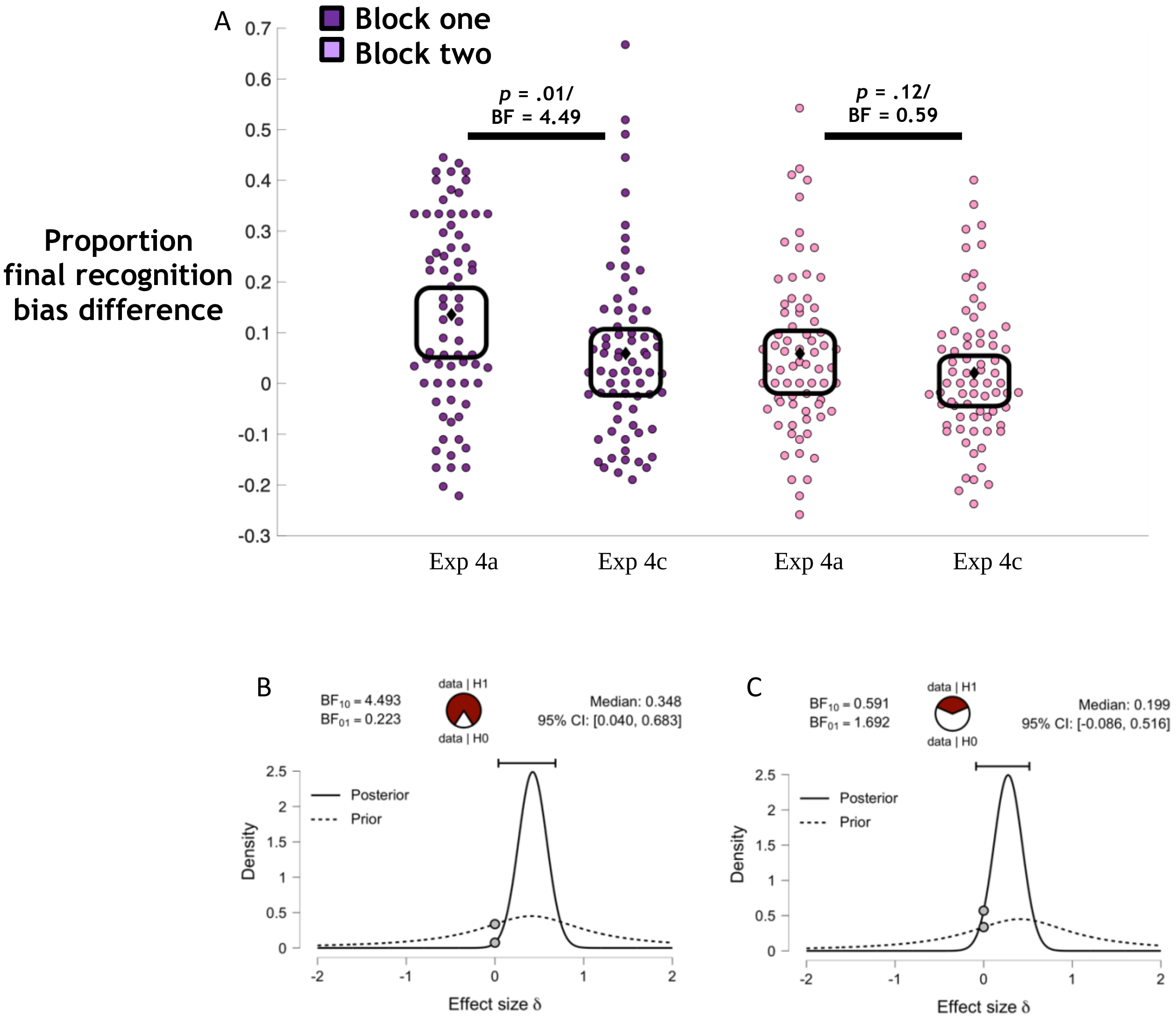

Figure 4.17. Final recognition bias difference from Experiment 4a and 4c. Within block one, the bias difference was larger following active retrieval versus select refresh (panel B). However, comparing experiments within block two showed no difference following either passive re-study conditions on the bias difference measure (panel C).

4.5. General discussion

The experiments presented in Chapter 4 aimed to test whether updating of face memories is specifically enhanced by active retrieval attempts or is instead related to subsequent encoding of interfering new faces irrespective of whether stored memories are retrieved. Across three experiments, a novel paradigm was designed to test how face recognition accuracy and bias on a final test differed based on whether a preceding refresh task required participants to retrieve or encode face memories. During final recognition tasks, retrieval accuracy measured whether the target face was correctly recognised, separately for memories shown during refresh (repeated trials) or not shown during refresh (baseline trials). In addition, and as a critical index of face memory updating, retrieval bias measured whether participants were more likely to repeat the same recognition errors from refresh to final recognition than to make new errors.

For final recognition accuracy, it was predicted that accuracy would be larger following refresh tasks requiring retrieval versus encoding of face memories, consistent with the retrieval practice effect literature (Butler et al., 2006; Marsh et al., 2007; Roediger & Butler, 2011; Roediger & Marsh, 2005). However, across all three experiments, measures of final recognition accuracy were not influenced by refresh task or trial type (repeated or baseline trials). Instead, the only effect found for final recognition accuracy showed that accuracy was significantly lower in block two than in block one, for all experiments, suggesting an order effect of block was present. Furthermore, final recognition bias measures were expected to be larger following refresh tasks requiring retrieval rather than encoding, consistent with the view that retrieval is a powerful mechanism for updating memory (Bridge & Voss, 2014; Roediger & Butler, 2011). Supporting this hypothesis, in Experiment 4a, final recognition bias was significantly higher following retrieval versus passive re-study of face memories. Experiment 4b established that final recognition bias was not significantly different between blocks that both included a retrieval task during refresh, suggesting that the bias measure was not influenced by order effects. Finally, Experiment 4c showed that final recognition bias was not enhanced if the refresh task required participants to select a face based on a distinctiveness judgement (and not based on memory) compared to passively re-studying face memories.

Consistent with views that retrieval is a powerful mechanism for modifying memories, across experiment analyses showed that, within block one trials, final recognition bias was enhanced following retrieval versus selection refresh tasks. This finding rules out a ‘self-choice’ effect to account for face memory updating (Murty et al., 2015; Watanabe, 2001; Watanabe & Soraci, 2004) and emphasises the importance of active retrieval attempts (‘retrieval-mode’) for the updating of episodic memories (Karpicke & Zaromb, 2010). For block two trials, it was shown that final recognition bias was enhanced following retrieval versus passive re-study refresh tasks, suggesting that active retrieval enhanced bias even when block order was kept constant. In sum, these results support the argument that retrieval versus re-study enhances the updating of memories, consistent with prior retrieval-induced updating findings (e.g. Bridge & Voss, 2014; Butler et al., 2006; Marsh et al., 2007; Roediger & Marsh, 2005).

In all analyses conducted within and across experiments, Bayes Factors were utilised in addition to traditional statistical analysis of null hypothesis significance testing. The strength of using Bayes Factors is emphasised within these experiments as a tool to assess whether the data from Experiments 4b and c were more consistent with one of two competing hypotheses. Specifically, several hypotheses were made to predict patterns of accuracy and bias during final recognition following refresh tasks (see Section 4.2.2, 4.3.2 & 4.4). With traditional inferential statistics such as null-hypothesis significance testing, the data could only be used to reject the null hypothesis. In contrast, Bayes Factors were able to provide the amount of evidence for one hypothesis versus another, including evidence in favour of the null hypothesis (Dienes 2011; 2016), providing more relevant statistical inferences in relation to the predictions made within the current chapter.

Focusing on final recognition bias measures, there are several potential explanations accounting for the enhancement of bias following retrieval versus re-study tasks. One suggestion, mentioned within the retrieval practice literature, is that retrieval leads to an elaboration of memory traces (Kornell et al., 2011; Roediger & Butler, 2011; Roediger & Karpicke, 2006). Elaboration of the memory trace following retrieval may be achieved through encoding of selected information during retrieval. Indeed, one suggestion made in previous chapters argued that recognition bias is mediated by increased encoding of faces during previous tasks. In the experiments of Chapter 4, it is possible that encoding of faces selected during refresh was enhanced during the retrieval tasks compared to re-study or selection. Increased encoding of selected information may have been induced by the difficulty of the retrieval vs. re-study task (e.g. Pyc & Rawson, 2009; Roediger & Butler, 2011). It could be argued that retrieval refresh task was more difficult due to the difficulty of retrieving and recognising face memories that shared perceptual characteristics, compared to re-study tasks that told participants which face to encode, or required participants to select faces on perceptual characteristics. It is also possible that the refresh tasks may have required different levels of attentional processing, for example, greater attention to faces in the retrieval compared to re-study tasks, which in turn would enhance the encoding of these faces for the subsequent final recognition task (Dudukovic, DuBrow & Wagner, 2009).

Alternatively, increased bias following retrieval vs. re-study tasks could consider the reactivation hypothesis that assumes the reactivation of memory traces is critical to induce the updating of memories (Gershmann et al., 2013; Hupbach et al., 2007; 2008; 2009; Sederberg et al., 2011) a key assumption of reconsolidation theory (Dudai &

Eisenberg, 2004; Hardt et al., 2010). For Experiments 4a-c, it is possible that the retrieval condition increased the reactivation of face memories, more so than in the re-study refresh tasks. The reactivation of target memories during retrieval vs. re-study may have made these memories more labile to modification of distractor faces selected during refresh. An alternative possibility is that distractor faces were reactivated during the retrieval refresh task, assuming that distractor faces have a representation in the brain, or in face space, as a result of sub-optimal encoding of target faces during the prior learning phase.

To resolve these competing suggestions, future work is needed that can measure the extent to which target and distractor face representations are reactivated during retrieval/re-study tasks. For example, advanced analytical tools of EEG data, such as pattern similarity techniques (including RSA; Kriegeskorte et al., 2008) have been developed to measure the extent to which neural activity from encoding tasks is reinstated during retrieval (Kuhl & Chun, 2014; Lee et al., 2018; Leiker & Johnson, 2014; Schapiro et al., 2018; Staresina et al., 2013; Thakral et al., 2017; Van den Honert et al., 2016). In this sense, if measures of reinstatement are observed during the recognition of target and/or distractor faces that are associated with biased future responses, this would support the argument that neural representations of face memories are reactivated and modified during recognition attempts.

Another consideration can question the assumption that reactivation of memories is specific to the active retrieval condition. On the contrary, reactivation may also be present during re-study tasks as reactivation can occur unintentionally after participants are reminded of encoding events (rather than as a result of actively trying to remember events), and such unintentional effects of reminders can induce memory updating (e.g. Hupbach et al., 2007, 2008, 2009). Reactivation has also been shown to occur spontaneously during periods of sleep (Schapiro et al., 2018) and wakeful rest (Schapiro et al., 2018; Staresina et al., 2013). It is therefore possible that reactivation of neural representations may occur to some extent during passive and select re-study tasks. Indeed, the final recognition bias estimates following the re-study tasks indicated that participants were more likely to select the same distractor, rather than switch to a different distractor (with the exception of the passive re-study task in Experiment 4c). Therefore, a reactivation mechanism may also have been present during re-study tasks, although the degree to which memories are reactivated during re-study would most likely be reduced compared to retrieval. These suggestions are only tentative, therefore future studies applying pattern similarity analyses of EEG data with the retrieval/re-study paradigm from Chapter 4 may determine the extent to which neural reactivation occurs within retrieval and re-study tasks. Measures of neural reinstatement would be expected to be lower during re-study if reactivation is a key mechanism that enhances updating following retrieval versus re-study.

The aforementioned arguments were presented to account for the updating of face memories where final recognition responses were biased towards faces selected during refresh. Thus, these theories account for the updating of memories by erroneous information that was inconsistent with original encoding. In contrast, memory for target faces during final recognition was not influenced by refresh task in any of the experiments from Chapter 4, contrary to the retrieval-practice hypothesis (whereby retrieved memories should be boosted following retrieval compared to re-study). However, the results are consistent with studies showing that retrieval may not benefit long-term retrieval accuracy compared to re-study when there is a short delay between initial tasks (i.e. refresh) and the criterial retrieval tasks (i.e. final recognition; Rowland & Delosh, 2014; Smith, Roediger & Karpicke, 2013; Wheeler, Ewers & Buaonanno, 2003). Additionally, retrieval practice benefits compared to re-study might only be expected when performance on the initial retrieval task is high, or when data are conditionalised on trials where correct information was selected during intermediate retrieval/re-study tasks (Rowland & Delosh, 2014; Smith, Roediger & Karpicke, 2013). In the present experiments, participants' selected the target face during refresh for 50% of trials in Experiment 4a, 46-56% of trials in Experiment 4b, and 32% of trials in Experiment 4c (when they were selecting faces based on distinctiveness rather than recognition). Therefore, memory for the correct target was only reinforced on a relatively low proportion of trials in the retrieval task, which may be why there was no difference in final recognition between retrieval and re-study refresh tasks. However, if the likelihood of target memory reinforcement at refresh is a key modulating factor of final recognition performance, this account would predict lowest performance after the distinctiveness task (Experiment 4c) than the other refresh tasks (in Experimens 4a-b), but no such lowering of performance was found. Instead, final recognition accuracy was very consistent across all three experiments.

What cannot be accounted for based on the reviewed retrieval practice literature is why final recognition accuracy for baseline trials was similar to that of repeated trials during final recognition. The majority of retrieval practice studies show improved accuracy for memories that have received retrieval practice (i.e. repeated trials) compared to memories that have not received retrieval practice (i.e. baseline trials; Roediger & Butler, 2011). However, the discrepant findings in Experiments 4a-c may be explained by substantial methodological differences used in these experiments compared to previous work. Nevertheless, focusing on studies that have a similar design to Experiments 4a-c, Smith et al. (2013) showed increased accuracy for items that received retrieval practice compared to items that did not receive retrieval practice even with a delay between study and test phases of 15 minutes (similar to the time lag between learning and final recognition phases for experiments in the current chapter which was roughly 20 minutes).

The absence of final recognition accuracy differences between repeated and baseline trials may be the result of stimulus/task design differences between the current experimental design and previous literature. For example, the study phase from Smith et al. (2013) required participants to view words for 2000ms, without making any judgements of the words. In contrast, faces in the present experiments underwent extensive learning procedures (repeated learning cycles, 5000ms presentation time, characteristic ratings of faces). Furthermore, novel faces have substantially different characteristics in terms of semantic/verbal associations or memorability compared to words, and may be processed by different brain systems (Loffler et al., 2005; Kanwisher & Yovel, 2006; Kriegeskorte, 2007; Pitcher et al., 2009; 2011). Consequently, given such procedural and stimuli differences to prior studies, there are a number of potential reasons why the design of the present experiments may not show differences in final recognition accuracy for face memories in the repeated and baseline condition.

As the conclusions of the retrieval practice effect are dependent on how data are analysed (e.g. Rowland, & Delosh, 2014; Smith et al., 2013), it is important to justify why final recognition accuracy was measured using all trials for Experiments 4a-c. If data were conditionalised on trials where targets had been selected during refresh, then a task confound could emerge. That is, in the active retrieval conditions, trials where participants selected a target may reflect more strongly/accurately encoded memories that were more accessible and recognisable during refresh. In contrast, when participants were told to encode a target face during passive re-study, or selected a target face during the distinctiveness task, participants may not necessarily have recognised this face as a target image. Consequently, repeating target selections during final recognition would be more likely following active retrieval owing to the strength of the memory representations, rather than the functional differences between the type of refresh task. Hence the current experiments did not conditionalise final test performance based on refresh accuracy (as done in the previous experiments in Chapter 3), since doing so would be confounded when comparing different refresh tasks.

In conclusion, the present experiments provide support for a retrieval-related mechanism that enhances updating of face memories. These findings advance previous knowledge in the area of episodic memory updating by establishing novel evidence for the updating of item memories as a result of recognition attempts. However, based only on this behavioural evidence, it is not possible to determine the specific neurocognitive mechanisms that underlie recognition-induced updating of face memories. This issue was addressed in subsequent experiments by using EEG methods in conjunction with the current paradigm, as presented in the next chapters.

Chapter 5 - ERP correlates of objective, subjective and updating processes during face memory retrieval

In Experiments 3a and 3b, it was found that high confidence recognition responses during initial retrieval predicted that participants would select the same face during subsequent retrieval, regardless of whether the initial recognition choice was correct or incorrect. To account for these findings, it was suggested that participants were encoding faces selected during initial retrieval that biased future retrieval attempts, at the expense of original target face memories. However, an alternative suggestion proposed that reactivation of face memories during Test 1, enhanced for high confidence retrieval judgements, modulated subsequent memory updating. Furthermore, the results from Chapter 4 established that updating of face memories is enhanced following active retrieval attempts versus re-studying of face memories, supporting the argument that the observed updating effects in Chapter 3 were retrieval-induced. Nevertheless, these arguments are only tentative based on behavioural data alone. One approach to gain a more detailed understanding of the neurocognitive mechanisms of face memory updating is to use EEG with the repeated recognition paradigm used in Experiment 3b. This approach, detailed in this chapter, allowed the investigation of ERP effects during repeated recognition that may index encoding versus reactivation during updating of face memories.

As reviewed in the introductory chapter, the ERP technique has been widely used to study the neurocognitive mechanisms of episodic memory encoding (e.g. Paller & Wagner, 2002) and retrieval (e.g. Rugg & Curran, 2007). More recently, ERPs have also been used to investigate mechanisms of retrieval-induced updating. In an initial study, Bridge and Paller (2012) found that during repeated cued-recall of object-location associations, ERPs during initial retrieval distinguished accurate from erroneous retrieval attempts from 400-700ms at frontal electrodes. In contrast, ERPs during initial retrieval that predicted whether participants would make a similar response in a subsequent recall task (i.e. updating) were found from 700-1000ms at frontal and parietal electrodes. Similar ERP findings were demonstrated by Liu et al. (2018) where participants recalled word pair associations in three separate tasks. Successful retrieval of word pairs was associated with ERP positivities from 400-700ms, whilst ERPs that predicted subsequent recall accuracy were found from 700-1000ms. The conclusion from these studies was that reactivation of episodic memories (in the earlier time window) was followed by the engagement of encoding mechanisms (in the later time window) that modified and updated memories.

Although Bridge and Paller (2012) and Liu et al. (2018) showed ERP evidence of the neural mechanisms underlying retrieval-induced updating, it is unknown whether similar mechanisms are engaged for simpler forms of retrieval, specifically, during the recognition of face memories. To investigate this issue, the ERP study in Chapter 5 was conducted to assess the ERP correlates of retrieval-induced updating of face memories. ERPs were analysed during both initial and subsequent retrieval attempts to delineate the ERP correlates of retrieval-induced updating as a function of accurate and erroneous retrieval attempts during initial recognition, and the consequences of initial recognition for subsequent remembering.

In the present EEG Experiment 5, a modified version of the paradigm from Experiment 3b was used. Participants initially encoded a number of face targets, and next completed several recognition trials in Test 1. Here, target faces were presented alongside distractor face images from the corresponding face set. Participants were asked to select the face they recognised from the learning task, as well as provide a confidence response during recognition. The same recognition task was completed during Test 2, for face sets that were shown in Test 1 (repeated trials) and face sets that were not shown during Test 1 (baseline trials). This repeated recognition design was used to examine the neural correlates of objective and subjective recognition, as well as updating-related mechanisms during both initial (Test 1) and subsequent (Test 2) retrieval tasks. To this end, ERPs were compared during both Test 1 and Test 2 according to objective recognition accuracy (correct target selection versus incorrect distractor selection). ERPs were also compared based on recognition confidence (high recognition confidence versus low recognition confidence) during both tests, separately for correct and incorrect recognition attempts. In both tests, ERPs were also contrasted as a function of the type of recognition response made in the other test. During Test 1, ERPs were compared based on whether participants later made a similar response in Test 2 ("future close bias", since participants selected a face in Test 2 that was close in image space to the face selected in Test 1) or made a dissimilar response in Test 2 ("future far bias", since participants selected a face in Test 2 that was far in image space to the face selected in Test 1). For Test 2 trials, ERPs contrasted neural activity when participants repeated a similar response to that in Test 1 (repetition close bias) or made a dissimilar response to that in Test 1 (repetition far bias).

In Test 1, it was predicted that encoding-related ERP activity would be present for recognition responses that participants would repeat in Test 2, consistent with Bridge and Paller (2012) and Liu et al. (2018). Such activity was predicted to be separable from ERP effects associated with objective recognition accuracy and subjective recognition confidence, since the latter contrasts should primarily reveal ERP markers of memory reactivation. In Test 2, reactivation-related ERPs would be expected when recognition responses were similar to those in Test 1, even when Test 2 recognition was erroneous, providing evidence that those errors were the results of face memories having become updated by previous retrieval attempts.

5.1. Methods

Participant

Forty-seven participants (37 female, M_{age} = 22.05, SD_{age} = 4.08) ranging in age from 18-35 were recruited from the University of Kent in a within-subjects design. This sample size was determined to maximise participant numbers in a time-limited recruitment period. Participant numbers were also determined to have an equal number of participants for each of the six counter-balanced versions of the experiment, (7 participants in each version), with 5 additional participants recruited to replace 5 participants during recruitment who provided noisy EEG data, to ensure a full counterbalancing. During data cleaning and pre-processing of EEG data, seven participants were excluded (three due to experimental error, four due to poor quality of EEG recordings). Of the eligible participants (N=40), there was 87% power to detect a medium effect size (0.5) and 46% power to detect a small effect size (0.3) with a two-tailed alpha of .05. Participants were eligible for participation if aged 18-35, were right-handed, had normal/corrected-to-normal vision and were neurologically healthy and not under any psychoactive medication. Participants were treated in line with ethical guidelines and procedures stated by the University of Kent School of Psychology ethics board. Participants received either cash or course credit in return for participation.

Stimuli

Face stimuli. As in previous studies, artificial face space models were used to create face stimuli, constructed of 29 components. From this model, 160 face sets were created with each set containing 9 face images. Initially, 160 target locations were sampled at an equal distance from the space origin, ranging from ~5,000 to ~15,000 Euclidean points in space (see Figure 5.1). Ten face sets were removed from the batch due to image artefacts.

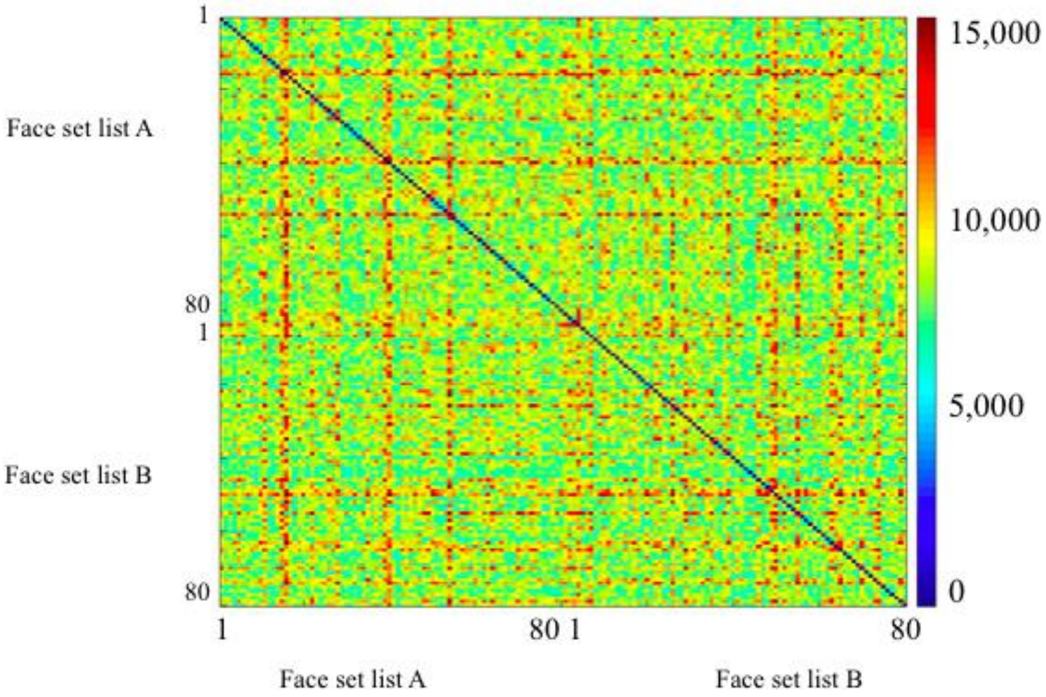

Figure 5.1. Euclidean distance between initial face locations. Colour map representing the Euclidean distances between the original 160 face locations in an artificial face space. Distances ranged from Euclidean distances of 5,000-15,000, ensuring perceptual dissimilarity between these face images.

For each of the 150 face locations, four additional faces were then randomly sampled from a uniform distribution limited to 1000-6000 points around each target face location. The first 75 face sets were assigned to face set list A, and 75 ‘matched’ face sets assigned to face set list B. Face set lists were matched to equal the pairwise Euclidean distances between the nine images within matched face sets. For example, Figure 5.2 shows face set two from list A and list B. This figure demonstrates that the pairwise

Euclidean distances are identical in face set two from list A and B. Importantly, face sets were perceptually dissimilar to one another, ensuring no overlap of image similarity across sets. As participants completed two separate EEG sessions (see procedure), controlling face set lists in this way removed any confounding influence of face stimuli differences in producing varied recognition measures between EEG sessions.

For the face recognition task, one of the five faces within a set was designated as the experimental target face shown during learning and recognition tests. Each of the five images across sets was selected as the experimental target an equal number of times. The four remaining faces from each set were designated as 'distractor' faces shown only during recognition tests. Face stimuli were counter-balanced according to whether face sets were presented during Test 1 and Test 2 (repeated trials) or were shown only during Test 2 (baseline trials). Additionally, face set lists were counter-balanced to be presented in either session one or two.

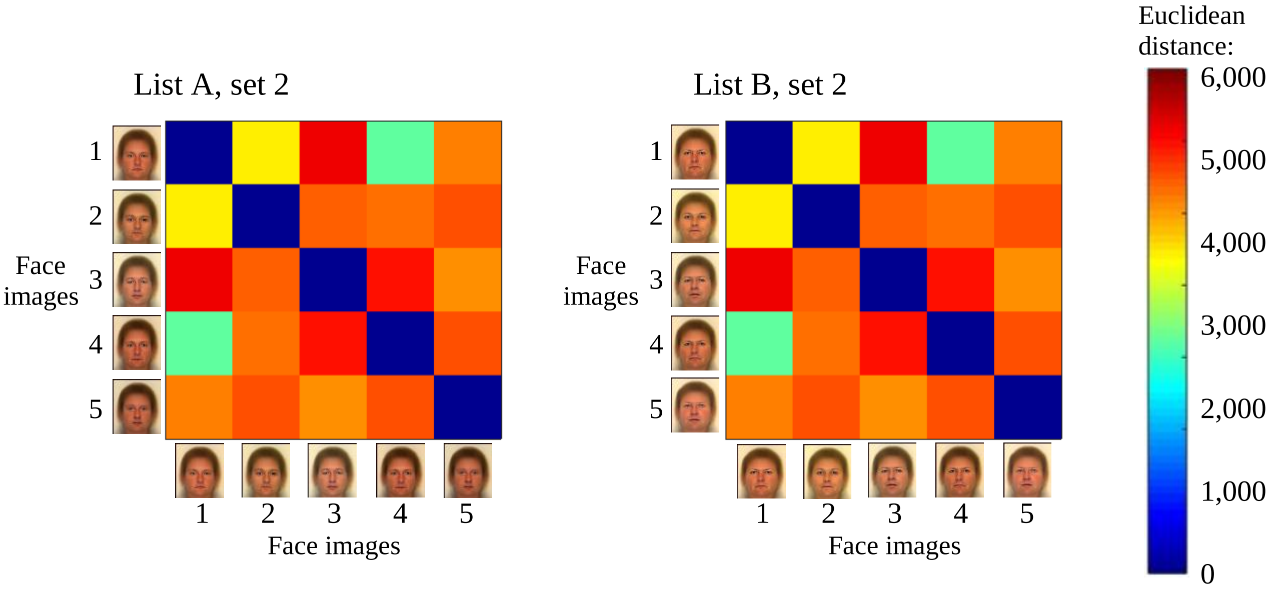

Figure 5.2. Euclidean distance between faces within sets. Colour maps are shown for face images in set two (list A top left, list B bottom left). Face sets were matched according to the pairwise Euclidean distances between the five face images within the pairs of sets.

Apparatus, procedure and design

The apparatus and procedure used in the present EEG experiment was almost identical to Experiment 3b (see section 3.2.1). The only task differences for the EEG experiment was the length and number of blocks, and a shorter learning task. In total, participants completed 10 blocks, split across two separate EEG sessions. Each block started with a learning phase where 15 target faces were encoded with two learning cycles (compared to three cycles used in previous chapters, see Appendix A). Following a short letter search filler task, participants completed 10 recognition trials in Test 1. For each trial, a target face was shown amongst four distractor faces, with participants required to select the face they recognised from the previous learning phase and provide a judgement of how sure they were that the target face was chosen by holding down the chosen key for longer or shorter (longer key presses indexing higher confidence). Following a second letter search filler task, participants completed the same recognition task in Test 2, for 10 repeated trials (shown during Test 1) and 5 baseline trials (not shown during Test 1). Five blocks were completed during session one, and after the fifth block in session one, participants were told to return to the EEG lab at the same time the following day to complete session two.

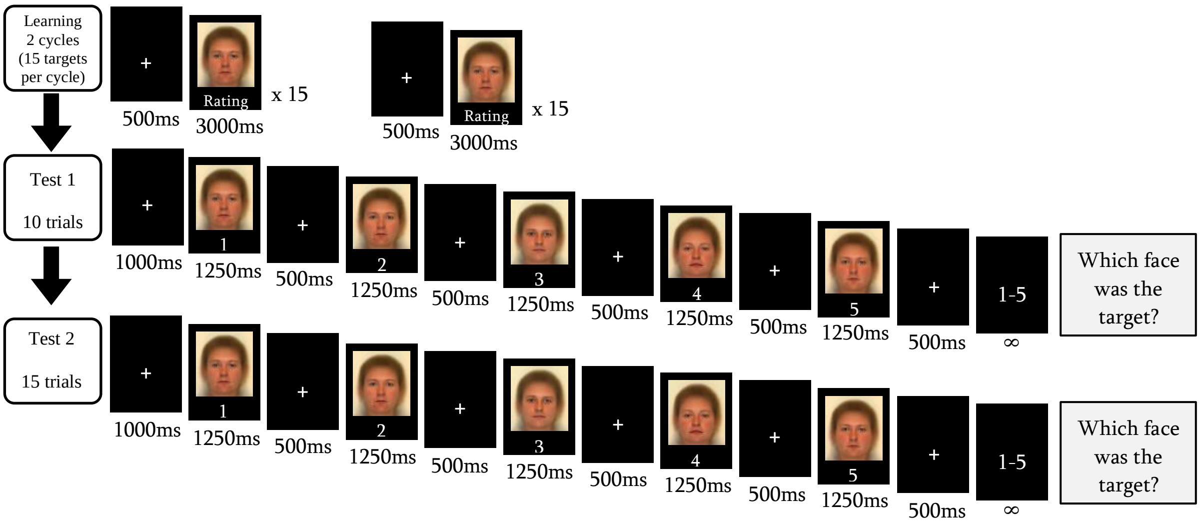

Figure 5.3. Procedure for Experiment 5. Participants encoded 15 face targets during a learning phase, following which participants completed two recognition tests, with 10 trials completed in Test 1 and 15 trials completed in Test 2 (included 10 ‘repeated’ trials and 5 ‘baseline’ trials. Five blocks were completed overall.

Electrophysiology

EEG was recorded during all phases with FCz as an on-line reference and a sampling rate of 500Hz. Sixty-four ‘active’ Ag/AgCl electrodes were affixed to an actiCap (Brain Products GmbH) based on the extended 10/20 system (Jasper, 1958), which was placed on the participants’ scalp. Electrode impedance with active electrodes was kept below 25kΩ (given that active electrodes amplify the EEG signal from the electrode, such impedance values are deemed acceptable for reliable EEG measurement). A Brainvision Quickamp (resolution/unit = 0.0715µV, amplifier low cut-off = DC, amplifier high cut-off = 140Hz) with an actiCAP control box was used to amplify and digitise recorded electrode signals. On-line filter cut-off points were set at 0.05 and 70Hz. Eye movements and blinks were recorded with VEOG placed below the left eye and HEOG placed on the right outer canthi. One electrode was fixed to the left and right mastoid each, with off-line re-referencing to the average mastoid EEG.

Off-line EEG data was analysed using EEGLAB (Delorme & Makeig, 2004), an open-access toolbox extension in MATLAB (MathWorks, Inc, Natick, MA). Pre-processing of EEG data was initially conducted on session one and session two data separately to avoid correcting artefacts that were specific to one session within the other session (e.g. electrodes that needed to be interpolated for session one but not session two). Recorded EEG data was first high-pass filtered (filter cut-off = 0.1Hz). EEG was then segmented into epochs of 2.1s (-0.5s to 1.6s) time-locked to face image onset for images selected during the Test 1 and 2, regardless of whether the selected image was the target or not (that is, only the ERPs for the one selected face out of the five shown faces in each trial was analysed). Excessively noisy segments of the epoched data (due to muscle activity, slow-drifts, etc) were removed prior to running an Independent Component Analysis using the runica function from EEGLAB (default extended-mode training parameters; Delorme & Makeig, 2004). Components associated with eye blinks/movements and high-frequency noise (muscle or faulty electrode noise) were manually identified using scalp map topographies, frequency power spectra and component activations across time for each trial, with identified components subtracted from the EEG recording. From the cleaned EEG, noisy electrodes were interpolated (average number of interpolated electrodes across participants = 3%), before session two was appended to the start of session one for each participant. Appended EEG data sets were then low-pass filtered (filter edge = 40Hz), with each epoch baseline corrected from -0.2 to 0s pre-stimulus. A final inspection of EEG manually identified and removed any remaining epochs that demonstrated excessive noise, prior to statistical analysis. Across participants, an average of 5% of epochs were removed.

Data analysis

Behavioural analyses. Participant’s behavioural responses were first analysed to replicate the results from Experiment 3b (see Section 3.2.2) and examine whether recognition performance differed with the changes in task for the EEG experiment. As a reminder, behavioural data were initially analysed to compare how recognition accuracy (proportion target selections) and confidence (mean keypress duration) differed for Test 1, Test 2 repeated and Test 2 baseline conditions. Furthermore, for each of these three conditions, the relationship between recognition accuracy (Euclidean distance between target and selected faces) and confidence (keypress durations) were also calculated (see Appendix E.1 for results). In addition, bias measures were calculated to examine how repeated recognition updated face memories. During Test 2, proportion bias (as well as bias difference measures) were compared following high compared to low confidence responses, separately for accurate and erroneous recognition attempts, during Test 1. High and low confidence responses were created by splitting trials according to the median confidence value for correct and incorrect Test 1 trials. This procedure was validated by comparing mean confidence values between correct high and low, and incorrect high and low confidence conditions (see Appendix E.3). Finally, the relationship between recognition accuracy during Test 1 (Euclidean distance between target and selected faces) and recognition bias during Test 2 (Euclidean distance between selected faces at Test 1 and 2) was examined to see if participants were more biased during Test 2 when previous responses were more similar (i.e. accurate) to target faces. All 40 participants who completed the EEG experiment were eligible for these behavioural analyses[1].

---

[1] As in previous chapters, participants were also excluded from bias analyses if they had accuracy rates of 0.2 or lower during test 1. However all 40 participants passed this criterion.

ERP analyses. The repeated recognition paradigm used in the present experiment allowed the analysis of ERP activity in both recognition tests. ERPs in Test 1 contrasted recognition accuracy (targets vs. distractor faces), confidence (high vs. low responses), and 'future bias', the latter referring to whether participants made similar or dissimilar recognition responses during future Test 2 recognition. Similarly, ERPs in Test 2 contrasted recognition, accuracy, confidence and 'repetition bias', referring to whether participants responses for repeated trials in Test 2 were similar or dissimilar to previous Test 1 recognition choices. Across all Test 1 and Test 2 conditions, participants were excluded if any of the ERP conditions had fewer than 10 trials, resulting in a final analysis sample of 30 participants[2]. A trial cut-off of 10 is lower than typical, recommended ERP trial numbers (Boudewyn et al., 2018). However, a 10 trial cut-off was decided to avoid losing a considerable number of participants which would have reduced the power of statistical analyses. Furthermore, trial-wise ERPs were inspected for each participant to ensure that any excessively noisy individual ERPs were removed to avoid these individual trials distorting the averaged ERP (which was more likely given the relaxed trial number cut-off).

Recognition accuracy and confidence. In the first set of ERP analyses, during both Tests 1 and 2, ERP activity was categorised as a function of recognition accuracy (correct or incorrect) and confidence (high or low). For accuracy, Test 1 and repeated Test 2 trials were categorised as to whether the target face was selected (correct) or whether any distractor face was selected (incorrect). The duration of keypresses were acquired for all trials, with longer keypresses indicating a higher confidence response. For correct trials, the median keypress length was calculated within participants, and trials with a keypress

---

[2] All 30 participants scored higher than 0.2 target recognition accuracy during test 1, meaning all participants showed memory for target faces.

length longer than the median assigned to the ‘correct high’ condition, and trials with a keypress length shorter than the median assigned to the ‘correct low’ condition. The same procedure was done for incorrect trials to produce ‘incorrect high’ and ‘incorrect low’ conditions. For Test 1, average trial numbers from the 30 eligible participants were 23 (Test 1 correct high), 23 (Test 1 correct low), 24 (Test 1 incorrect high) and 24 (Test 1 incorrect low). For Test 2, average trial numbers were 21 (Test 2 correct high), 21 (Test 2 correct low), 27 (Test 2 incorrect high) and 27 (Test 2 incorrect low).

Recognition accuracy and bias. In a second, separate set of ERP analyses, Test 1 and 2 trials were also categorised according to recognition accuracy and future bias (in Test 1) and accuracy and repetition bias (in Test 2). For Test 1, trials were initially categorised according to whether the target face was selected (correct) or whether a distractor was selected (incorrect). For both trial types, the Euclidean distance between faces selected during Test 1 and faces selected during the matched Test 2 trial were acquired. This ‘future bias’ value thus represents the extent to which participants made similar (smaller future bias) or dissimilar (larger future bias) responses between recognition tests. For correct Test 1 trials, a median Euclidean future bias value was calculated within participants. Trials where the future bias value was below the median were assigned to the correct close future bias condition, whereas trials with future bias values above the median assigned to the correct far future bias condition. A similar approach was done for incorrect Test 1 trials, producing incorrect close and incorrect far future bias conditions[3].

A similar approach was taken for Test 2 trials. For repeated trials during Test 2, the

---

[3] Bias conditions were created using a median split of all responses in order to maximise signal-to-noise ratio of averaged ERPs. If ERP conditions were created from exact repetitions (similar to how behavioural data was analysed) then a significant proportion of participants would have been excluded from analyses due to insufficient trial numbers. In a supplementary analysis (not presented) ERPs were created using this method, with ERP conditions showing similar amplitudes to conditions created using the median split procedure.

Euclidean distance between faces selected during Test 2 and faces selected during the preceding matched Test 1 trial were acquired, providing a ‘repetition bias’ value. A similar median split procedure was used for Test 2 correct and incorrect trials to categorise trials according to whether Test 2 responses were similar (close repetition bias) or dissimilar (far repetition bias) to previous Test 1 recognition. For Test 1, averaged trial numbers were 28 (correct close future bias), 18 (correct far future bias), 25 (incorrect close future bias) and 24 (incorrect far future bias). For Test 2, averaged trial numbers were 28 (correct close repetition bias), 14 (correct far repetition bias), 27 (incorrect close repetition bias) and 27 (incorrect far repetition bias)[4].

Behavioural measures of accuracy, confidence and bias. For all ERP conditions described above, behavioural measures of recognition accuracy, confidence and bias were calculated (extracted from trials included in ERPs) in order to characterise behavioural performance across conditions. Recognition accuracy (a.k.a recognition “error”) referred to the mean Euclidean distance between selected face images and the target face from the corresponding face set (smaller values indicate better recognition accuracy). Recognition confidence was calculated as the mean keypress durations for all trials within a condition (larger values indicate higher recognition confidence). Finally, recognition bias was calculated as the mean Euclidean distance between the selected face in one test and the face selected in the matched trial from the alternate test (smaller values indicate increased recognition bias). Whilst some of these measures were trivial for certain conditions (e.g. all correct conditions will have a mean recognition accuracy of 0), descriptive statistics for these measures were used to facilitate the interpretation of reported ERP effects. Inferential

---

[4] Trial numbers for the correct far condition were lower than trial numbers in other conditions as a consequence of the median split of Euclidean bias values. For participants who had >50% repeated accurate responses across Test 1 and 2, the median Euclidean bias value will be 0 (as correct recognition responses have Euclidean values of 0). Therefore, any correct response during Test 1 for these participants will be automatically assigned to the correct close condition, allocating more trials to this condition and fewer trials to the correct far condition.

statistics between conditions for each measure were also performed, these are presented in Appendix E.4.

Cluster-based permutation ERP statistical analysis. For all ERP analyses, cluster-based permutation tests were performed with ERP amplitude (μV) as the dependent variable. Several repeated measures designs were conducted, separately for Test 1 accuracy and confidence, Test 1 accuracy and future bias, Test 2 accuracy and confidence, and Test 2 accuracy and repetition bias. Between test comparisons were also done for the accuracy and confidence conditions, with the between test comparison of accuracy and bias located in Appendix E.5 for interest[5]. For main effects, interaction and simple main effects within each of these analyses, paired samples t-tests (two tails, alpha threshold = .05) were performed between the two conditions of interest, with data samples at 60 electrodes (VEOG and HEOG excluded) included from 0-1.1s[6]. The first step of the cluster-based permutation tests required the calculation of paired t-test values at each electrode-time data sample. A t-test value that was lower than the critical alpha threshold (.05) was included in a cluster so long as the t-test for at least 2 neighbouring spatio-temporal samples was also lower than the alpha threshold. In the second step, cluster-level statistics were calculating by summing t-values for data samples included in the cluster. The size of clusters then underwent significance testing by creating a permutation null distribution (5000 resamples). The proportion of permuted t-values larger than the observed cluster-level statistic in this distribution provided the significance (*p*-value) of the cluster-level statistics (two-tails, cluster alpha threshold = .05). Clusters that had a *p*-value

---

[5] Between-test analyses of accuracy-bias were not part of the main analysis due to the difficulty in comparing bias measures between tests, as the interpretation differs for the bias measures calculated in Test 1 and 2.

[6] The analysed time window was shorter than the total face presentation time as errors in the PsychoPy scripts meant that a small number of face presentations were shortened by the next fixation cross being presented earlier than expected. Thus, time windows of 1.1s were analysed to avoid the risk analysing ERP activity that did not correspond to faces on-screen.

below the cluster alpha threshold were interpreted based on spatial and temporal dimensions, illustrated with line and topographical plots.

5.2. Results

Behavioural results: influence of repeated retrieval on objective and subjective recognition

Initially, accuracy and confidence were analysed separately for the three recognition test conditions (Test 1, Test 2 repeated, Test 2 baseline). As a reminder, results from Experiment 3b demonstrated that proportion accuracy was higher during Test 1 compared to Test 2 conditions, with no difference found between conditions for mean confidence.

Proportion accuracy. A one-way repeated measures ANOVA compared proportion accuracy rates between Test 1 (M = 0.52, SE = 0.02), Test 2 repeated (M = 0.44, SE = 0.02) and Test 2 baseline conditions (M = 0.40, SE = 0.02). The ANOVA was significant ($F(2, 78) = 45.46$, $p < .001$, $\eta^2p = 0.54$). Post-hoc comparisons (corrected $\alpha = .016$) demonstrated that accuracy was higher during Test 1 compared to both Test 2 repeated ($t(39) = 8.03$, $p < .001$, $g = 1.26$) and Test 2 baseline conditions ($t(39) = 8.37$, $p < .001$, $g = 1.32$). Furthermore, accuracy was significantly higher for the Test 2 repeated versus baseline condition ($t(39) = 2.73$, $p = .009$, $g = 0.43$). These results differ to those in Experiment 3b, where recognition accuracy between Test 2 conditions was not different.

Mean confidence. A second one-way repeated measures ANOVA compared mean confidence between Test 1 (M = 1.49, SE = 0.06), Test 2 repeated (M = 1.35, SE = 0.07) and Test 2 baseline conditions (M = 1.37, SE = 0.05). The ANOVA was not significant

($F(1.15,44.76) = 2.84$, $p = .09$, $\eta^2p = 0.07$), indicating no influence of initial, repeated or delayed recognition on average confidence responses, similar to the results from Experiment 3b.

Accuracy and confidence of Test 1 recognition modulates subsequent accuracy and updating

To examine behavioural measures of recognition bias from the present EEG experiment, recognition bias during Test 2 was analysed according to whether participants repeated the same recognition responses to those made in Test 1, separately for Test 1 recognition responses that were correct with high confidence, correct with low confidence, incorrect with high confidence and incorrect with low confidence. Confidence conditions were formed by categorising responses in Test 1 as to whether they were longer or shorter than participant-level median durations (see Appendix E.3 for validation of confidence median split). Similar to the experiments in Chapter 3, updating was operationalised as a proportion bias measure (proportion of Test 2 trials where the same face was selected from Test 1; see Figure 5.5 for descriptive statistics). Analysing this measure, a 2 (Test 1 accuracy; correct, incorrect) x 2 (Test 1 confidence; high, low) repeated measures ANOVA showed main effects of Test 1 accuracy ($F(1,39) = 151.68$, $p <.001$, $\eta^2p = 0.80$) and Test 1 confidence ($F(1,39) = 5.16$, $p = .03$, $\eta^2p = 0.12$) were observed. An interaction effect, however, was not significant ($F(1,39) = 0.41$, $p = .53$, $\eta^2p = 0.01$). Planned one-sided paired samples t-tests were performed to compare proportion bias between Test 1 correct high versus correct low conditions and Test 1 incorrect high versus incorrect low conditions (inferential statistics in Table 5.1). For correct Test 1 trials, there was no significant difference between high and low confidence, with the Bayes Factor for this t-test showing more support for the null hypothesis. However, bias was significantly higher for the Test 1 incorrect high confidence compared to low confidence condition, with the

Bayes Factor for this test providing more support for the alternative hypothesis (that bias would be larger for incorrect high vs. incorrect low conditions). These findings partially support the hypothesis that Test 1 confidence predicts Test 2 bias, conflicting with data from Experiments 3a and 3b where correct high confidence had larger bias than the correct low confidence condition.

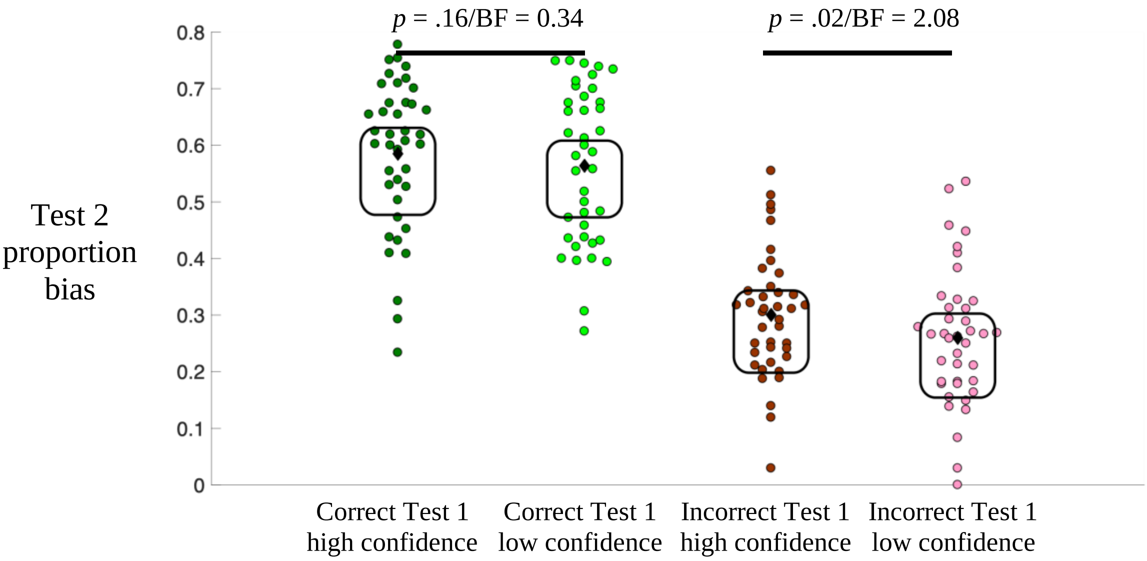

Figure 5.5. Proportion bias statistics for Experiment 5. Test 2 proportion bias for Test 1 correct high, correct low, incorrect high and incorrect low conditions.

Table 5.1. Inferential statistics for bias analysis in Experiment 5. Statistical results for the comparison of proportion bias and bias difference measures during Test 2, for correct high-low and incorrect high-low confidence contrasts.

High-low confidence constrast	df	*t*	*p*	*g*	B	Prior
Correct, proportion bias	39	1.01	.16	0.16	0.34	0.40
Incorrect, proportion bias	39	2.06	.02	0.33	2.08	0.31

Relationship between Test 1 accuracy and Test 2 bias. The final replication analysis assessed whether participants' accuracy during Test 1 correlated with biased recognition attempts during Test 2. That is, for each trial in the Test 1 and Test 2 repeated conditions, the Euclidean distance between selected faces and the target face during Test 1 was calculated (Test 1 error). Additionally, the Euclidean distance between the selected faces at Test 1 and Test 2 was acquired (i.e. Test 2 bias). Pearson correlations were computed between Test 1 error and Test 2 bias values, with the mean coefficient value significantly different than a point value of zero (*t*(39) = 13.69, *p* <.001; see Figure 5.6). A similar pattern of findings was shown when only analysing trials where a distractor face was selected during Test 1 (*t*(39) = 7.44, *p* <.001). Bayes Factors for both contrasts also provided more evidence in favour of the alternative vs. null hypothesis (Test 1 all trials; $BF_{+0(1.17,0.707)}$ = 1.47e +14); Test 1 incorrect trials – $BF_{+0(1.05,0.707)}$ = 9.46e +6). Similar to the results seen in Experiments 3a and b, the consistent positive correlation between Test 1 error and Test 2 bias indicates that recognition responses during Test 1 that were more similar to target faces led to more biased recognition responses during Test 2.

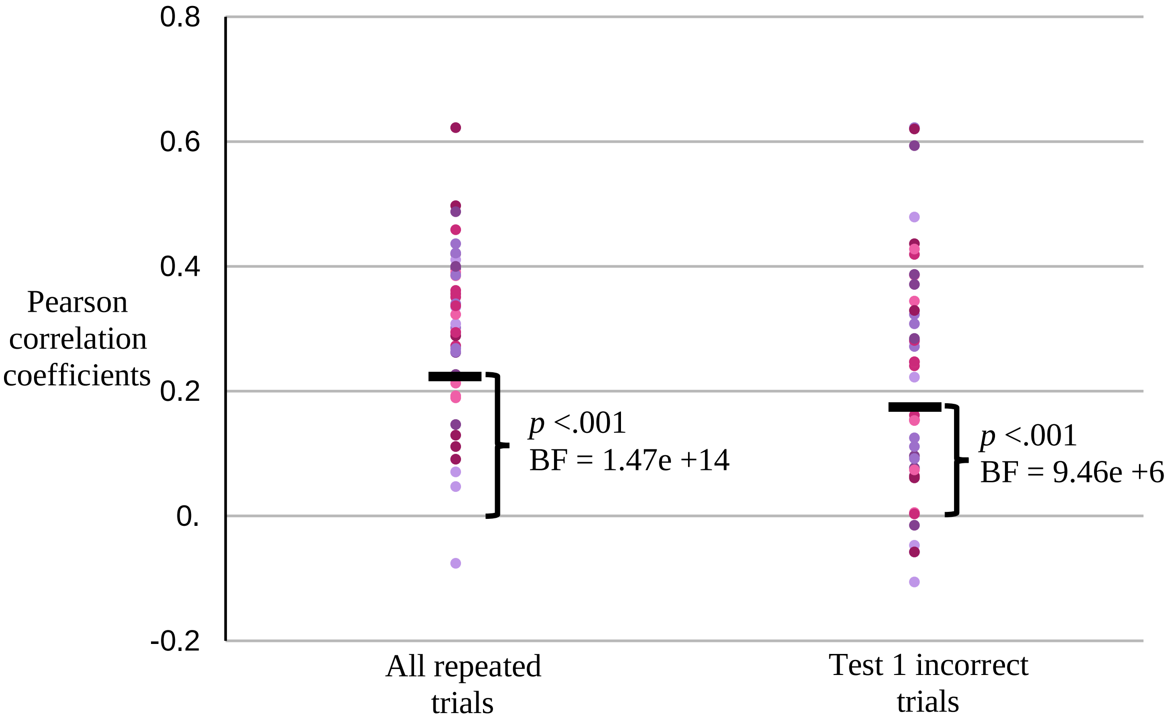

Figure 5.6. Correlation between Test 1 accuracy and Test 2 bias. Relationship between Test 1 Euclidean error and repeated Euclidean bias for all Test 1 trials and incorrect Test 1 trials only.

Behavioural measures of ERP conditions

The next analysis compared behavioural measures of recognition error, confidence and bias of the ERP conditions, only for participants who provided sufficient data for each ERP condition. These analyses were performed in order to provide a behavioural interpretation of the ERP analysis in the subsequent section, with data from all 30 participants eligible for ERP analysis.

ERP conditions of accuracy and confidence. Firstly, behavioural measures are shown for ERP conditions of correct high confidence, correct low confidence, incorrect high confidence and incorrect low confidence. Descriptive statistics for all analyses can be found in Table 5.2.

Recognition confidence. As seen in Table 5.2, and as expected from the median split of confidence responses, mean confidence was larger for high vs. low confidence conditions, in both tests, an effect that appeared within both correct and incorrect responses separately. In addition, mean confidence was larger for correct compared to incorrect trials in both Test 1 and 2.

Recognition error. Recognition error (measured as the Euclidean distance between selected and target faces) was zero for all correct conditions. For the incorrect conditions in Test 1 and 2, mean recognition error was larger than 4.0. As noted in Appendix B (see Table B.1), when face pairs have a Euclidean distance greater than ~3.5, faces appear to be perceptually distinct. This suggests that, on average, faces in the incorrect high and low confidence conditions were perceptually distinct from targets.

Recognition bias. Recognition bias, measured as the Euclidean distance between faces selected at Test 1 and 2 for each recognition trial, was smaller for correct vs. incorrect, and for high vs. low confidence conditions. This suggests that participants selected faces that were more similar to responses in the alternate test for targets compared to distractors, and for high rather than low confidence recognition judgements. For both correct and incorrect conditions, recognition bias was smaller for high vs. low confidence conditions.

Table 5.2. Behavioural descriptive statistics for accuracy and confidence ERP conditions. Mean error, bias and confidence for correct high, correct low, incorrect high and incorrect low conditions in Test 1 and 2. Standard error values are in parentheses. Cells that are blank are correct conditions where all recognition error values were 0.

	Test 1			Test 2		
Condition	Error	Bias	Confidence	Error	Bias	Confidence
Correct		1.50 (0.08)	1.85 (0.09)		1.82 (0.08)	1.64 (0.09)
Incorrect	4.25 (0.03)	3.00 (0.08)	1.10 (0.09)	4.29 (0.03)	3.17 (0.08)	1.17 (0.09)
High confidence	2.08 (0.02)	2.11 (0.08)	2.12 (0.09)	2.10 (0.02)	2.03 (0.08)	2.06 (0.09)
Low confidence	2.17 (0.02)	2.72 (0.08)	0.83 (0.09)	2.19 (0.02)	2.64 (0.08)	0.75 (0.09)
Correct high		1.43 (0.11)	2.57 (0.11)		1.08 (0.09)	2.37 (0.11)
Correct low		2.22 (0.09)	1.13 (0.08)		1.92 (0.13)	0.91 (0.09)
Incorrect high	4.17 (0.04)	2.78 (0.09)	1.67 (0.09)	4.21 (0.04)	2.99 (0.08)	1.76 (0.11)
Incorrect low	4.33 (0.04)	3.23 (0.09)	0.54 (0.08)	4.38 (0.04)	3.36 (0.07)	0.58 (0.08)

ERP conditions of accuracy and bias. Next, behavioural measures of recognition bias, error and confidence were assessed based on the ERP conditions of accuracy and bias (correct close, correct far, incorrect close and incorrect far) for both Tests 1 and 2. Descriptive statistics for all conditions can be found in Table 5.3.

Recognition bias. Measures of recognition bias were, as expected from the median split of bias values, smaller for ‘close’ compared to ‘far’ bias conditions, observed for both correct and incorrect responses in both Test 1 and 2. Furthermore, recognition bias was smaller for correct vs. incorrect trials, showing that recognition responses in one test were on average more similar to responses in the alternate test for selected targets vs. distractor faces.

Recognition error. Recognition error was again zero for all correct conditions. For the incorrect conditions in Test 1 and 2, mean recognition error was similar to or greater than 4.0, suggesting that faces in these conditions appear to be perceptually distinctive to target faces. Further, mean recognition error was smaller for close vs. far bias conditions, indicating that when participants were recognising similar faces between tests (i.e. close rather than far bias), these faces were more perceptually similar to target images.

Recognition confidence. Mean recognition confidence was larger for correct vs. incorrect conditions in both tests. Additionally, mean confidence was larger for close compared to far bias conditions, which applied to both correct and incorrect trials in both Test 1 and 2, suggesting participants were more confident when selecting faces that were more similar to recognition responses in an alternate test.

Table 5.3. Behavioural descriptive statistics for accuracy and bias ERP conditions. Mean accuracy, bias and confidence for correct close, correct far, incorrect close and incorrect far conditions during Test 1 and 2. Standard error values are in parentheses.

	Test 1			Test 2		
Condition	Error	Bias	Confidence	Error	Bias	Confidence
Correct		2.23 (0.06)	1.79 (0.08)		2.10 (0.06)	1.58 (0.08)
Incorrect	4.26 (0.03)	3.03 (0.06)	1.09 (0.08)	4.30 (0.03)	3.19 (0.06)	1.18 (0.08)
Close bias	1.97 (0.02)	0.73 (0.06)	1.58 (0.08)	1.96 (0.02)	0.80 (0.06)	1.51 (0.08)
Far bias	2.28 (0.02)	4.49 (0.06)	1.31 (0.08)	2.33 (0.02)	4.53 (0.06)	1.28 (0.08)
Correct close		0.12 (0.05)	1.99 (0.09)		0.04 (0.02)	1.80 (0.09)
Correct far		4.34 (0.05)	1.59 (0.09)		4.16 (0.06)	1.36 (0.10)
Incorrect close	3.94 (0.05)	1.34 (0.12)	1.17 (0.08)	3.93 (0.05)	1.55 (0.11)	1.23 (0.09)
Incorrect far	4.57 (0.03)	4.73 (0.03)	1.02 (0.08)	4.66 (0.02)	4.82 (0.03)	1.12 (0.09)

ERP cluster analysis

The next analysis compared ERP activity according to recognition accuracy and confidence, as well as accuracy and bias, in both tests 1 and 2. Comparisons between tests were also performed to assess the consequences of repeated recognition on ERPs corresponding to recognition accuracy, confidence and updating.

ERPs of accuracy and confidence. Cluster-corrected permutation t-tests were performed to compare ERP activity for conditions of correct high confidence, correct low confidence, incorrect high confidence and incorrect low confidence responses. ERP differences between conditions were thus examined to assess main effects of recognition accuracy (correct versus incorrect), confidence (high versus low) and an accuracy by confidence interaction (see Table 5.4 for cluster statistics of both Test 1 and 2). As a reminder, and consistent with previous ERP research (Bridge & Paller, 2012; Liu et al., 2018; Rugg & Curran, 2007; Wilding & Ranganath, 2011; Woodruff et al., 2006; Woroch & Gonalves, 2010), it was predicted that ERPs for correct trials were expected to be more positive than incorrect trials. Similarly, ERPs for high confidence trials were predicted to be more positive than ERPs for low confidence trials.

Test 1. Firstly, in Test 1, a significant positive cluster were found for the main effect of accuracy, with ERPs for correctly selected target faces being more positive than ERPs for incorrectly selected distractors (Table 5.4). Figure 5.7 shows that this cluster corresponded to differences from ~0.4-1.1s, including electrodes across the whole scalp. In addition, two significant clusters were found for the main effect of confidence. These clusters show that ERPs for faces selected with high confidence were more positive than faces selected with low confidence from approximately 0.6-0.9s (cluster one) and 1.0-1.1s (cluster two) across central and posterior electrodes. The absence of any significant clusters for the interaction term suggests that ERPs distinguished high from low confidence conditions for both correct and incorrect Test 1 trials, however there was a non-significant trend towards an interaction. Separate tests of the effects of confidence for correct versus incorrect trials (Table 5.4) showed that target faces that were selected with high confidence were associated with a broadly distributed and sustained positive ERP effect compared to target faces selected with low confidence. However, ERPs to distractor faces incorrectly selected with high confidence were only more positive than ERPs to distractor faces selected with low confidence for a briefer duration and only across posterior electrodes (see Figure 5.8), and this cluster was not statistically significant.

Table 5.4. Cluster statistics for main effects and interactions of Test 1 accuracy and confidence. All significant clusters are reported. For effects with no significant clusters, the largest cluster (across positive and negative clusters) for each contrast is reported.

	Test 1		Test 2	
Cluster contrast	Cluster *t*	*p*	Cluster *t*	*p*
Correct-incorrect	45234.22	<.001	37946.85 3368.79	<.001 .04
High-low confidence	22159.51 3136.81	.002 .04	34522.02	<.001
Interaction	2114.87	.07	5710.88	.02
Correct high-correct low	22017.86	<.001	29091.22 3603.67	.001 .04
Incorrect high-incorrect low	1769.10	.08	806.37	.21

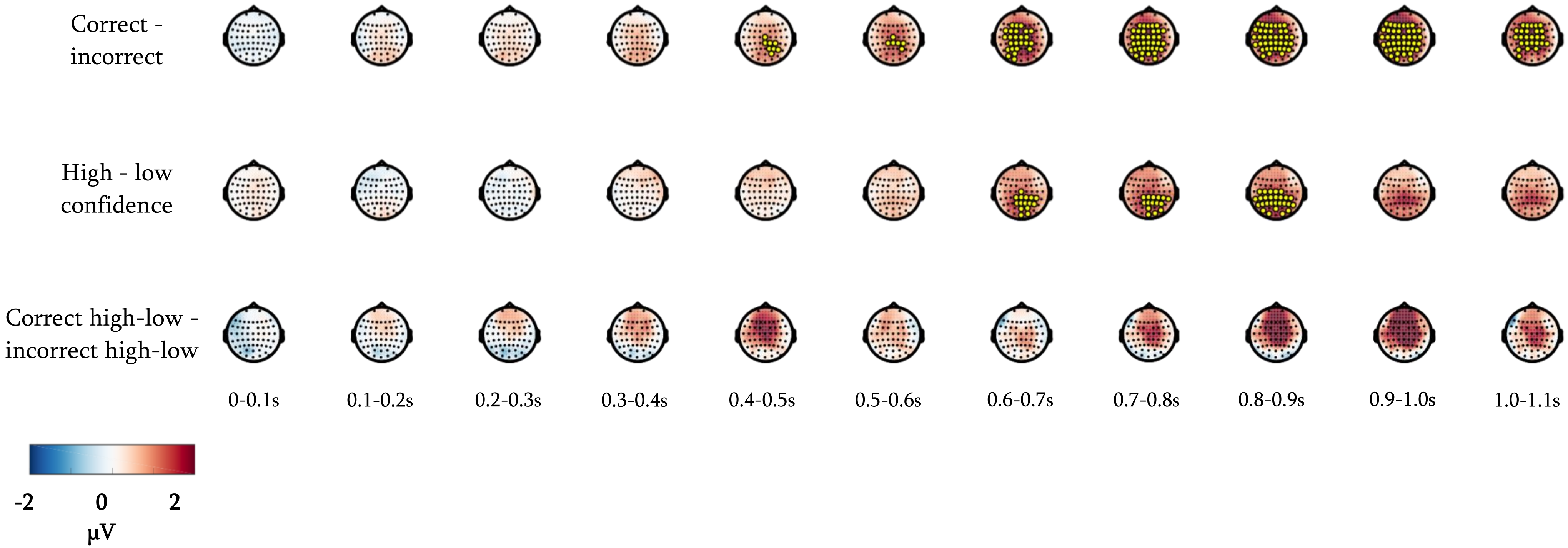

Figure 5.7. Cluster plots for ERP analysis of Test 1 accuracy and confidence. Topographical maps with cluster statistic results showing ERP differences for main effects of accuracy and confidence, as well as an accuracy x confidence interaction, in Test 1. Yellow dots denote when and where a significant cluster was found. Only the largest cluster is plotted.

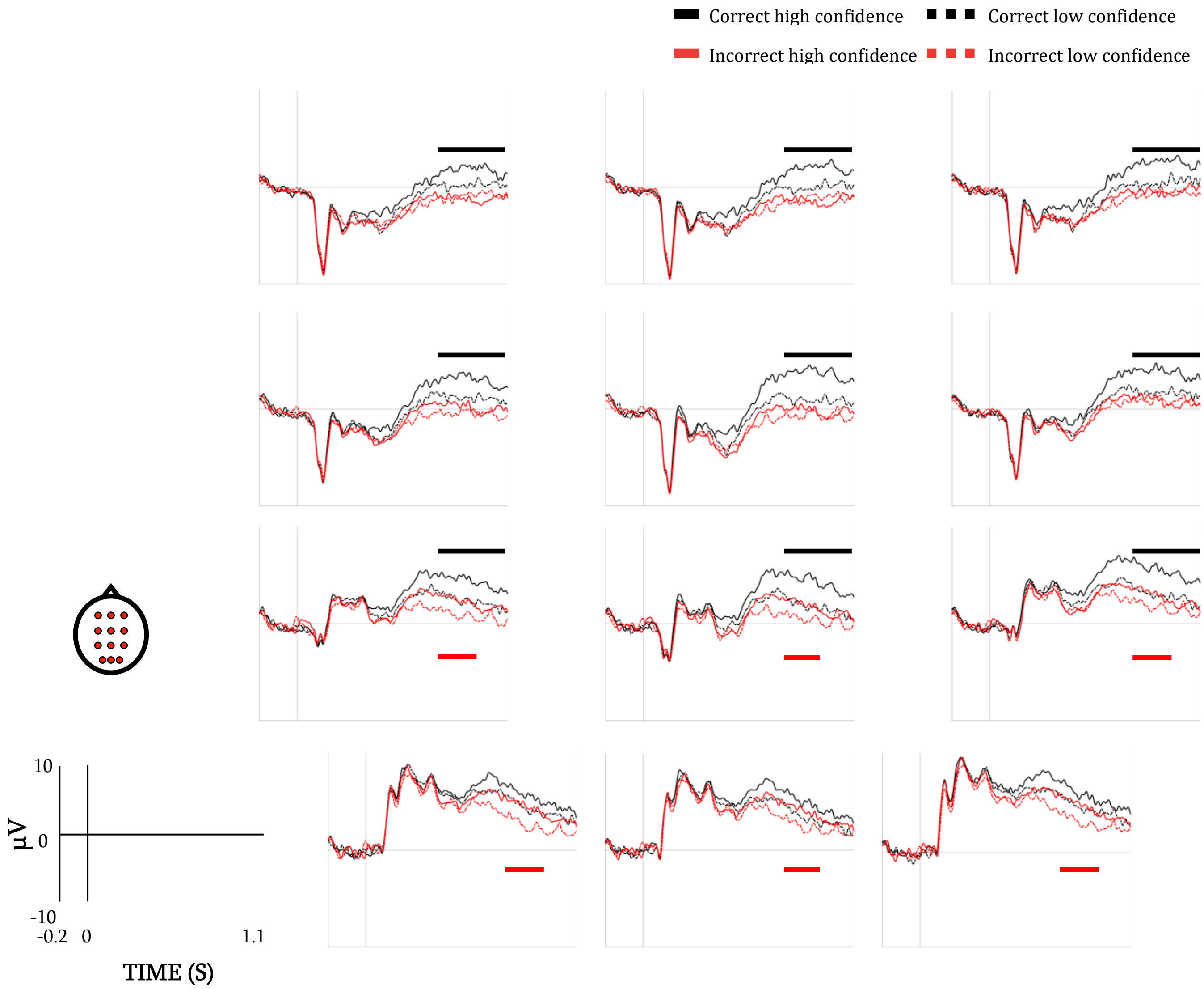

Figure 5.8. Grand-average waveforms for Test 1 conditions of accuracy and confidence. Waveforms plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Black and red lines denote significant times and electrodes for the accuracy and confidence main effect, respectively.

Test 2. Next, ERPs were compared as a function of accuracy and confidence within Test 2. Significant positive clusters were established for both the main effects of accuracy and confidence. As noted in Figure 5.6, both of these clusters were seen to include electrodes from across the whole scalp from 0.7-1.1s for the correct-incorrect contrast, and

0.6-1.1s (cluster one) and 0.4-0.5s (cluster two) for the high-low confidence contrast. Similar to Test 1, these findings indicate that ERPs were more positive for selected target faces vs. selected distractor faces, as well as faces selected with high vs. low confidence. In contrast to Test 1, the cluster for the interaction between Test 2 accuracy and confidence was significant. Follow-up cluster tests (see Table 5.4) were conducted separately within correct and incorrect trials, with two clusters for the simple main effect of correct high-correct low confidence seen to be significant. These clusters corresponded to more positive ERPs for target faces selected with high confidence compared to low confidence from approximately 0.6-1.0s (cluster one) and 0.4-0.5s (cluster two) at centro-posterior electrodes (see Figure 5.10). In contrast, the cluster for the simple main effect of confidence for incorrectly selected distractor faces high was not significant. Thus, analysis of Test 2 showed more positive ERPs for high confidence versus low confidence judgements only when target faces were correctly recognised, and not when distractor faces were incorrectly chosen.

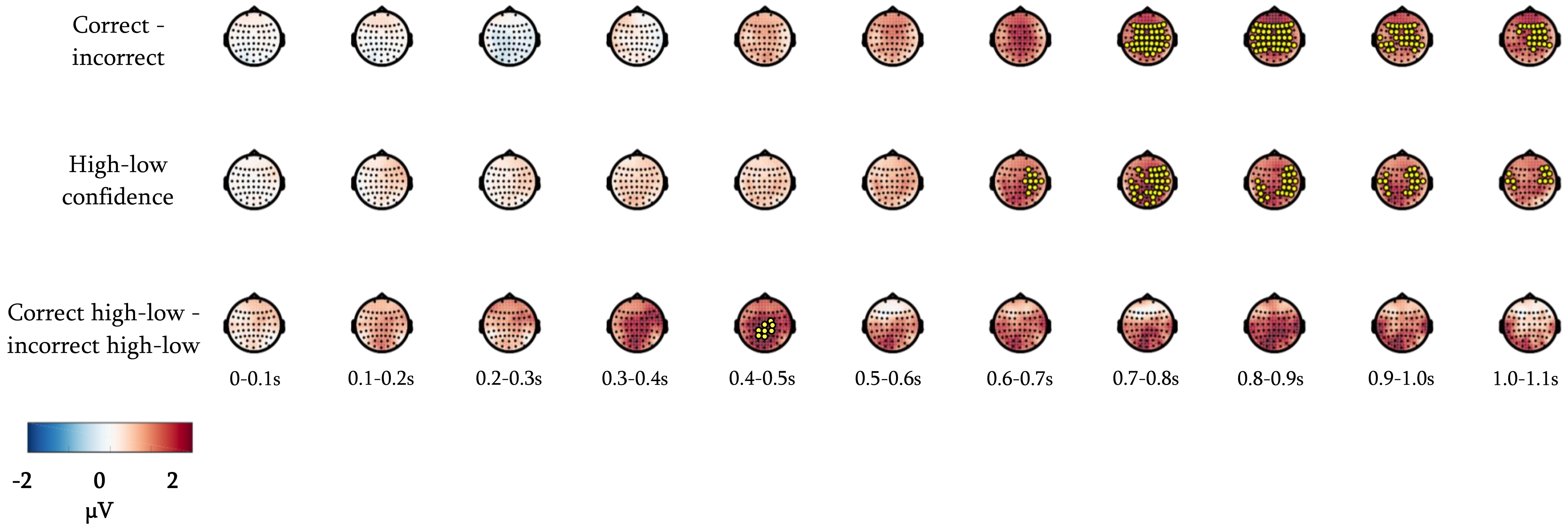

Figure 5.9. Cluster plots for ERP analysis of Test 2 accuracy and confidence. Topographical maps with cluster statistic results showing ERP differences for main effects of accuracy and confidence, as well as an accuracy x confidence interaction, in Test 2. Yellow dots denote when and where a significant cluster was found. Only the largest cluster is plotted.

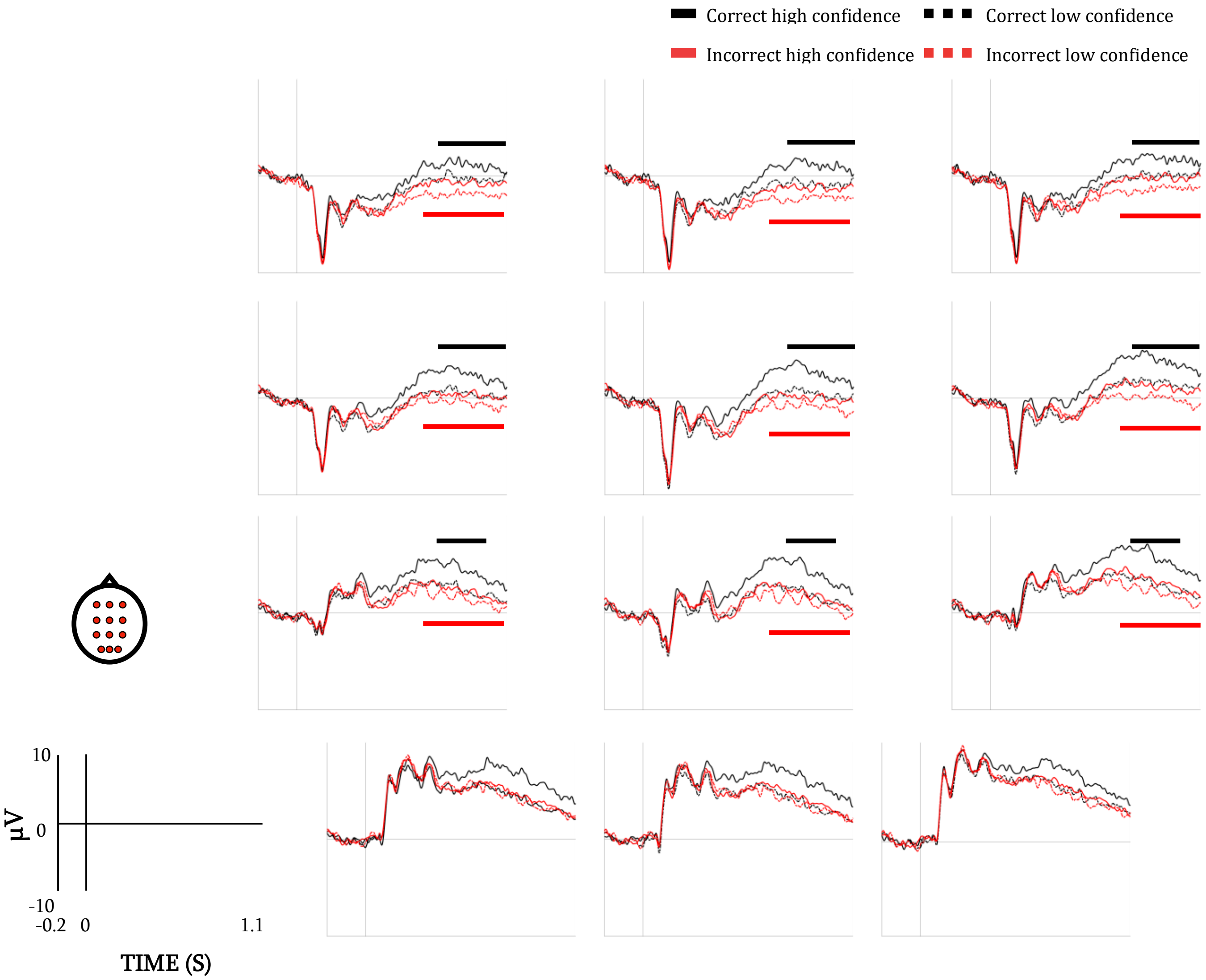

Figure 5.10. Grand-average waveforms for Test 2 conditions of accuracy and confidence. Waveforms plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Black and red lines denote significant times and electrodes in clusters for the main effects of accuracy and confidence, respectively.

Test 1 vs Test 2. Final ERP analysis of accuracy and confidence determined whether ERPs of these conditions statistically differed between tests 1 and 2. For this analysis, cluster corrected paired samples t-tests were performed between tests 1 and 2, with dependent variables of the accuracy-confidence interaction (i.e. correct high-correct low confidence difference minus incorrect high-low confidence difference) as a three-way interaction term. Two-way interactions also compared the difference between correct and incorrect ERP conditions, and high confidence and low confidence ERP conditions, between tests. Finally, ERPs for the correct high, correct low, incorrect high and incorrect low confidence conditions were compared between Test 1 and 2. As seen in Table 5.5, no significant clusters for these effects were found, suggesting that all ERP effects were similar between Test 1 and 2.

Table 5.5. Cluster statistics comparing Test 1 versus Test 2 for accuracy and confidence conditions. All significant clusters are reported. For effects with no significant clusters, the largest cluster (across positive and negative clusters) for each contrast is reported.

Cluster contrast	Cluster *t*	*p*
Accuracy x Confidence x Test	-534.41	.31
Accuracy x Test	669.89	.25
Confidence x Test	-118.19	.54
Correct high-low difference	-363.55	.38
Incorrect high-low difference	50.97	.64

ERPs of accuracy and bias. Next, ERPs during Test 1 and 2 were compared, using cluster-corrected permutation t-tests, according to recognition accuracy (correct versus incorrect), recognition bias (close bias versus far bias), and an accuracy-bias interaction (see Table 5.6 for cluster statistics). As a reminder, the bias measure in Test 1 separated ERPs for faces that were selected based on whether participants made a similar response in Test 2 (future close bias) versus a dissimilar response in Test 2 (future far bias). During Test 2, ERPs to selected faces were analysed based on whether participants selected a similar face to that they had selected previously in Test 1 (close repetition bias) versus a dissimilar face to that selected previously in Test 1 (far repetition bias). During Test 1, it was anticipated that ERP activity that differentiated close versus far bias conditions would reflect encoding mechanisms during retrieval that led to making a similar recognition response during Test 2. Such an encoding mechanism should be similar for both correct and incorrect Test 1 recognition. In Test 2, ERPs were expected to differentiate the close and far repetition bias conditions for both correct and incorrect Test 2 responses, providing evidence of the retrieval of updated face memories.

Test 1. As seen in Tables 5.6 and Figure 5.11, a positive cluster for the main effect of accuracy was significant, showing that ERPs for correctly selected target faces were more positive than ERPs for incorrectly selected distractor faces (as would be expected since this contrast was identical with that presented in the previous accuracy x confidence analysis). Additionally, analysis of ERPs in Test 1 showed that the cluster for the main effect of future bias was just below the cluster alpha threshold. This cluster corresponded to time points between 0.7-0.8s post-stimulus, showing that ERPs were more positive in the future close versus future far bias condition for less than 0.1s of the time window at posterior sites (though no cluster points can be seen Figure 5.8 as the cluster did not pass over at least one 100ms time boundary). As the size of this cluster was very small, this cluster may reflect random noise between the future close and future far bias conditions. Furthermore, the cluster for the interaction effect was not significant. In summary, ERPs during Test 1 predominantly reflected recognition accuracy, with minimal evidence found of ERP markers corresponding to memory updating found (see Figure 5.12).

Table 5.6. Cluster statistics for main effects and interactions of accuracy and bias. All significant clusters are reported. For effects with no significant clusters, the largest cluster (across positive and negative clusters) for each contrast is reported.

	Test 1		Test 2	
Cluster contrast	Cluster *t*	*p*	Cluster *t*	*p*
Correct-incorrect[1]	45234.22	<.001	37946.85	<.001
Close-far bias	2566.00	.05	2035.56	.06
Interaction	773.07	.21	4726.36	.03
Correct close-correct far	1524.35	.10	16976.20	.002
Incorrect close-incorrect far	1093.58	.26	-335.91	.40

[1] The correct-incorrect contrast is identical to that seen in Table 5.5 because the same trials are allocated to the correct and incorrect conditions for both analyses.

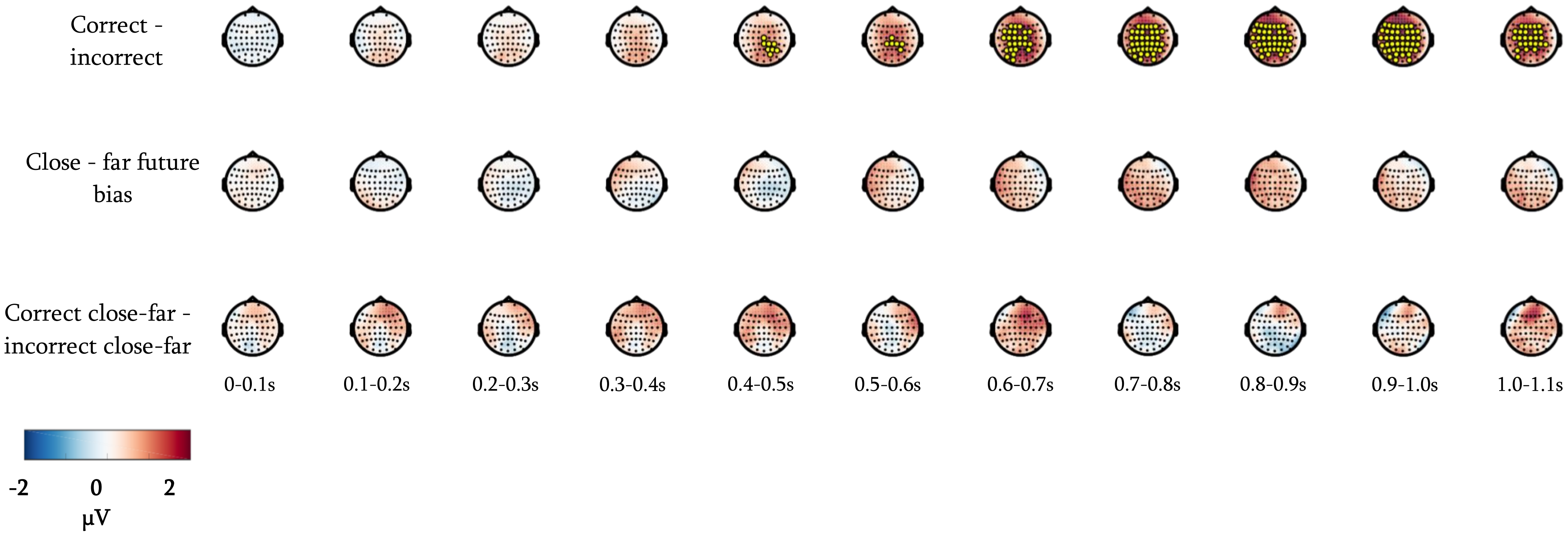

Figure 5.11. Cluster plots for ERP analysis of Test 1 accuracy and bias. Topographical maps with cluster statistic results showing ERP differences for main effects of accuracy and future bias, as well as an accuracy x bias interaction, in Test 1. Yellow dots denote when and where a significant cluster was found. Only the largest cluster is plotted.

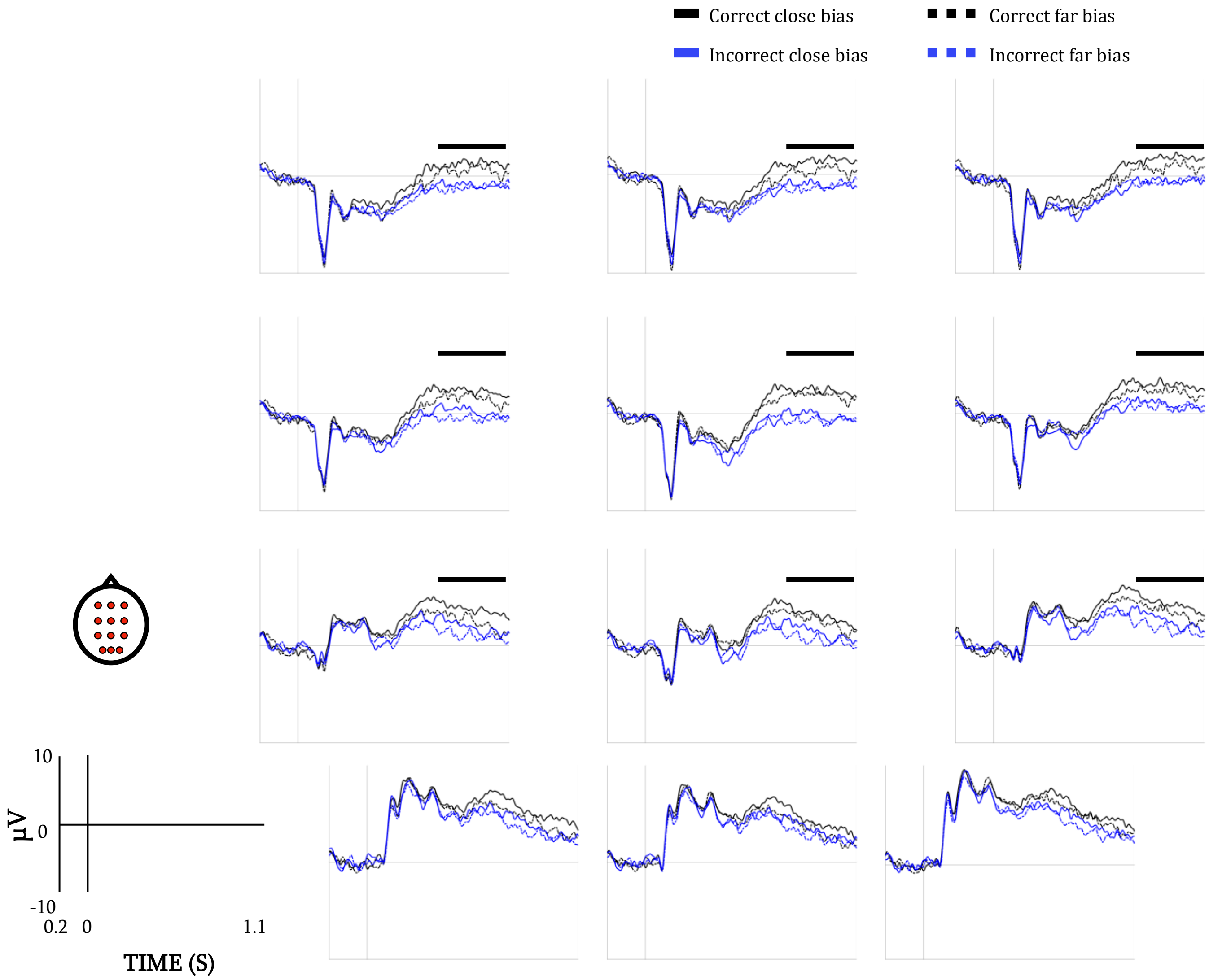

Figure 5.12. Grand-average waveforms for Test 1 conditions of accuracy and bias. Waveforms plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Black lines denote significant times and electrodes for the accuracy main effect.

Test 2. For Test 2, a significant positive cluster demonstrated that ERPs for correctly selected target faces were again more positive than ERPs for the incorrectly selected distractor faces (see Table 5.6 & Figure 5.13). In contrast, the cluster size for the main effect of repetition bias was not significant. However, a significant cluster was found for the interaction between Test 2 accuracy and repetition bias. Follow-up analysis showed a significant cluster for the simple main effect of correct close versus correct far repetition bias. Waveform plots in Figure 5.14 showed that this cluster corresponded to ERP positivities for correctly selected target faces during Test 2 for face sets where participants had also selected a target, or face similar to the target on Test 1, when compared to correctly selected targets for sets where participants had previously made a dissimilar error on Test 1. This positivity lasted approximately from 0.5-0.8s across most of the scalp. The cluster for the simple main effect of incorrect close versus incorrect far repetition bias, however, was not significant. To summarise, ERPs in Test 2 were more positive for correct recognition of target face memories, but only for recognition of targets that were similar to previous recognition attempts.

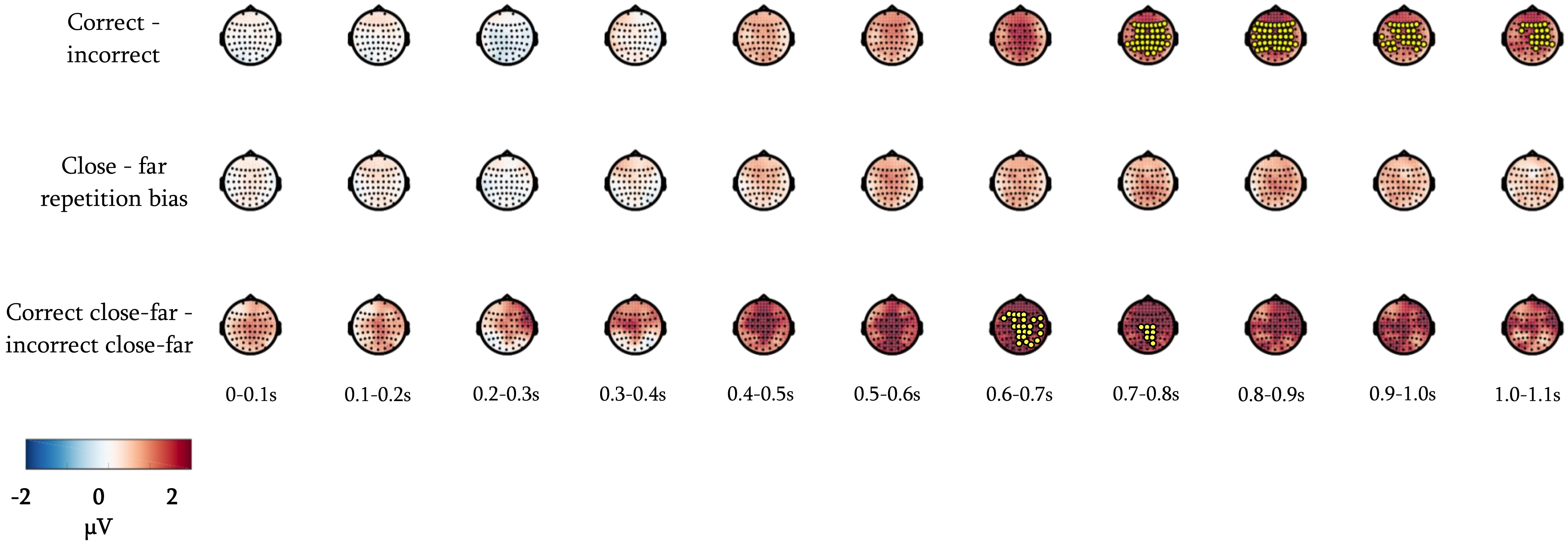

Figure 5.13. Cluster plots for ERP analysis of Test 2 accuracy and bias. Topographical maps with cluster statistic results showing ERP differences for main effects of accuracy and future bias, as well as an accuracy x bias interaction, in Test 2. Yellow dots denote when and where a significant cluster was found. Only the largest cluster is plotted.

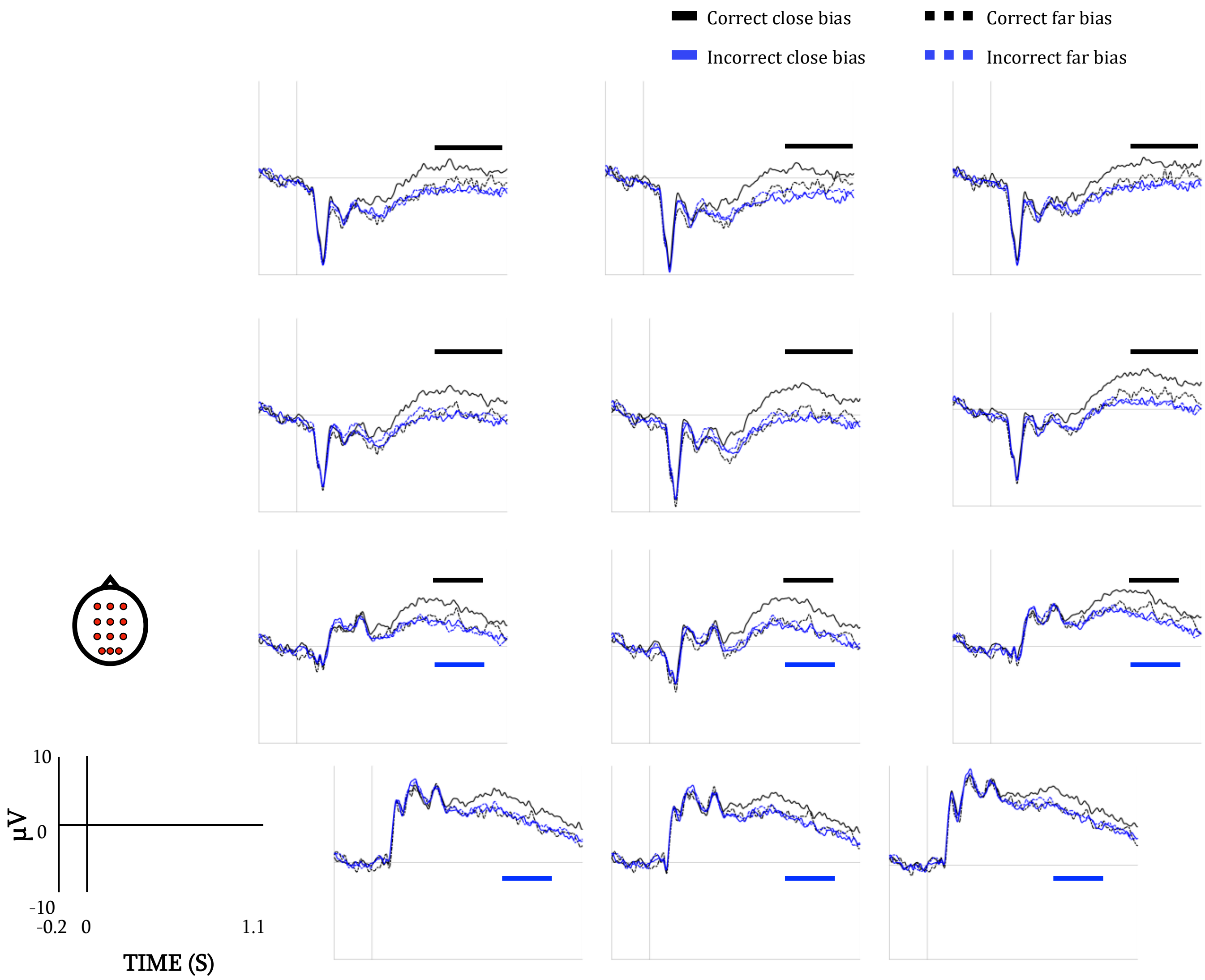

Figure 5.14. Grand-average waveforms for Test 2 conditions of accuracy and bias. Waveforms plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Black and blue lines denote significant times and electrodes for the accuracy main effect and correct simple main effect, respectively.

5.3. Discussion

The goal of the research in the present chapter was to investigate the ERP correlates of objective, subjective and updating processes during repeated face memory retrieval.

During initial (Test 1) and subsequent (Test 2) retrieval attempts, ERPs were more positive for accurately selected target faces than inaccurately selected distractor faces, an effect widespread across the whole scalp and sustained from ~400ms after stimulus presentation. Similarly, high versus low confidence recognition during Test 1 was related to increased ERP positivities, predominantly at centro-posterior sites, for both correct and erroneous recognition attempts. A similar effect was found during Test 2, however the ERP confidence effect was only apparent for correct recognition trials. Finally, ERPs during Test 1 did not differentiate whether participants would be biased or not during subsequent Test 2. In contrast, during Test 2, ERPs were more positive for correct recognition attempts when participants had recognised a similar face during the previous recognition test.

The ERP effect shown during Test 1 and 2 for the correct recognition of target faces can be interpreted as a marker of successful memory retrieval. These findings are consistent with previous studies showing that ERPs are more positive for successful episodic memory retrieval (Rugg & Curran, 2007; Wilding & Ranganath, 2011), including successful retrieval of face memories (Mackenzie et al. 2018; Mackenzie & Donaldson, 2007; Mackenzie & Donaldson, 2009). In the present study, ERP differences for correct versus incorrect recognition were sustained from around 0.4-1.1s and were widespread across the scalp. Interestingly, this ERP positivity associated with recognition accuracy appeared to be larger during Test 1 compared to Test 2, however these differences were not statistically significant between test meaning these results must be interpreted with caution. Possibly, ERPs were more similar between correct and incorrect recognition during Test 2 due to recognition of distractor faces from the previous test. It is important to emphasise that, during Test 2, incorrect recognition choices were, on average, perceptually dissimilar to target faces. Consequently, it is plausible to assume that distractor faces were recognised during Test 2 because these distractor images were encoded during the preceding Test 1 recognition task, as opposed to being more perceptually similar to target faces in Test 2 vs. Test 1.

In addition to objective accuracy effects, the analysis also showed ERPs associated with elevated subjective confidence in recognition. During Test 1, ERPs were more positive for recognition responses made with high compared to low confidence, for both correct and incorrect Test 1 responses, the timing and distribution of this effect lasting from 0.6-0.9s at central and posterior electrodes similar to previous ERP studies of recognition confidence (Addante et al., 2012; Woodruff et al., 2006; Woroch & Gonsalves, 2010). The ERP positivity for high compared to low confidence conditions in Experiment 5 applied to both correct and erroneous Test 1 recognition attempts (though the latter contrast did not meet the significance threshold). The fact that ERPs for erroneous recognition attempts were distinguished by confidence is interesting when considering that, for both Test 1 incorrect high and incorrect low conditions, the average Euclidean distance between targets and recognition errors was significantly greater than the distance at which faces become perceptually dissimilar. These results therefore indicate this ERP effect reflects a neural correlate of subjective retrieval mechanisms, irrespective of whether the recognised face resembled original face memories.

The finding that ERPs separated the recognition of novel distractor faces with high confidence can be linked by propositions from face space models (Lewis, 2004; Valentine et al., 2015). Specifically, it might be expected that, for participants to have a subjective appraisal during recognition of novel faces, these faces may be represented in face space and in the brain. As mentioned in the discussion section of Chapter 3 (see Section 3.3), distractor faces that had never been perceived prior to Test 1 may have a representation due to faulty encoding of target faces during the learning phase (assuming that the location of distractors within face space is within an ‘error boundary’ of a target face). These arguments were also proposed in line with signal detection theories of recognition (Busey et al, 2000; Parks & Yonelinas, 2009; Thakral et al, 2015; Wixted & Mickes 2010) and face space model assumptions (Lewis, 2004), whereby the strength of recognition is determinant on the signal match generated between perceived information during recognition and mnemonic representations stored in the brain. However, these 'global strength' arguments of recognition would be expected to modulate correlates of familiarity-based recognition judgements, such as the FN400 (Rugg & Curran, 2007; Wilding & Ranganath, 2011), or posterior ERP effects from ~300-500ms related to familiarity judgements for unknown faces (e.g., Mackenzie & Donaldson, 2007). Instead, the ERP results in this chapter demonstrated widespread positivities associated with the recognition of distractor faces, when these memories were associated with high confidence judgements. These ERP results are more consistent with recollection ERP correlates, with such effects seen across the whole scalp including parietal electrodes (similar to the LPE; Rugg & Curran, 2007; Wilding & Ranganath, 2011) and more anterior electrode locations (similar to Mackenzie et al., 2018; Mackenzie & Donaldson, 2009).

The centro-posterior location of the ERP confidence effect may suggest that the neural generator of these ERPs originate in posterior brain regions. However this is a very tentative suggestion as the neural source of ERP data was not assessed in this chapter. Previous work has linked activity in the precuneus with increased subjective vividness ratings during retrieval (Richter et al., 2016), although Bergström, Henson, Taylor & Simons (2013) source-localised a sustained negative rather than positive-going ERP to the precuneus, associated with successful recollection. Additionally, posterior brain regions may also be recruited in order to reactivate neural representations of retrieved memories, with neural representations of faces possibly reactivated in memory-related regions such as the angular gyrus (Lee & Kuhl, 2016), or face-specific brain regions including fusiform face gyrus (Loffler et al., 2005; Kanwisher & Yovel, 2006; Kriegeskorte, 2007) and occipital face area (Pitcher et al., 2009; Pitcher et al., 2011). However, the ERPs that have been associated with these face-specific regions, such as the N170 and N250, occur much earlier than the findings from the current chapter and generally refer to the detection or recognition of a previously seen face. Nevertheless, it may still be possible that these regions are implicated in subjective decisions during retrieval (e.g. Wiese et al., 2019).

During Test 2, ERPs were more positive for high confidence versus low confidence recognition, only when participants correctly selected target faces during Test 2, with no difference found when distractors were selected during Test 2. The absence of an ERP difference between high and low confidence conditions for incorrect trials during Test 2 may again be explained by the fact that, during the second test, participants were more likely to recognise distractor faces as being from the previous recognition test. Such recognition would be less likely to occur in Test 1 when distractor faces were completely novel to participants. Furthermore, when interpreting the grand average waveforms of incorrect high and low confidence conditions in Test 1 and 2, it appears that ERPs are more positive for the incorrect low confidence condition in Test 2 vs. Test 1 (whereas the ERPs for incorrect high confidence trials appear to be similar across tests). These differences are not statistically reliable, however the descriptive statistics support the idea that participants may have recognised distractors in the low confidence condition in Test 2. It is important to add that participants would only recognise distractor faces from the prior test based on a familiarity of having seen these faces from Test 1, rather than assuming that these images were targets from learning (given that participants were not more confident for faces in the incorrect low condition in Test 2 vs. Test 1). In contrast, the presence of the confidence effect for correct Test 2 trials, located predominantly at centro-posterior electrode sites again, suggests elevated confidence during retrieval may be driven by reactivation of neural representations of vivid face memories in regions critical for face processing.

One goal of the present chapter was examine whether some ERP effects during repeated recognition were indicative of the updating of face memories. During Test 1, a minimal ERP positivity was shown during recognition of memories that would be repeated during Test 2. However, the timing of this effect lasted less than 100ms, therefore this finding most likely reflects spurious ERP differences between the future close and future far bias conditions. However, during Test 2, it was shown that ERPs were more positive during the accurate recognition of face memories, only when participants had also recognised a similar face in the preceding Test 1 task. Given that the timing and topography of these ERPs were similar to those for accurate retrieval of memories associated with high confidence, and given that participants were more confident that faces in the correct close conditions were target faces, this ERP positivity most likely corresponds to retrieval processes associated with reactivation of strong memory representations. What remains unknown, however, is whether these face memories were strengthened by successful retrieval of these memories in Test 1, or whether retrieval and reactivation of these memories occurs due to strengthened encoding of these memories during initial learning.

These findings are somewhat similar to ERP effects in previous research studying the ERP correlates of retrieval-induced updating (Bridge & Paller, 2012; Liu et al., 2018). In both of these studies, ERP positivities were found during an initial retrieval test that predicted similar responses during subsequent retrieval. However, these studies reported dissociable effects of retrieval success (from 400-700ms) and re-encoding of the retrieved information that predicted future memory updating (700-1000ms). The data from the present research is partially consistent with these studies in that, during Test 1, ERPs were more positive for memories that were correctly recognised. On the other hand, no ERP effects were found during Test 1 that predicted the updating of memories during subsequent retrieval, as evidenced by the lack of significant clusters for the Test 1 bias analysis. A further similarity from the ERP results of the present chapter can also be made to the Liu et al. (2018) study. In their study, ERPs during a second test were more positive for memories that had already been successfully retrieved during Test 1, compared to memories that were correctly retrieved during Test 1 but incorrectly retrieved during Test 2. These findings are similar to the present data, where ERPs were more positive for recognition of target memories during Test 2 when participants recognised a similar face during Test 1. Thus, this ERP effect can be seen as reflecting the retrieval of memories strengthened by successful repeated accurate recognition.

It is interesting that the present results share similarities to the studies by Bridge and Paller (2012) and Liu et al. (2018) when considering the differences between type of episodic memories tested, and the retrieval tasks used between these studies. The face images used in the present chapter contained minimal semantic or conceptual information. On the other hand, stimuli commonly used in the episodic memory literature, such as from Bridge and Paller (2012) and Liu et al. (2018), use words or objects that are loaded with semantic/conceptual information. Previous work has established that ERP correlates of episodic memory retrieval may vary according to the semantic nature of tested memories (Mackenzie et al., 2018; Mackenzie & Donaldson, 2007; 2009; Nie et al., 2014), yet the data from the present chapter, compared to previous work studying the ERP correlates of retrieval-induced updating, share some similarities.

Furthermore, the type of retrieval task used in the present study, that being recognition of item memories, differs to previous studies of episodic memory updating that used cued-recall tasks. Previous work has demonstrated that the simpler and complex forms of episodic memory retrieval differ in brain regions recruited (Ranganath & Ritchey, 2012). Additionally, the amount of neural reactivation during retrieval, a critical factor for memory updating (Bridge & Paller, 2012; St. Jacques et al., 2013), correlates with the complexity/difficulty of retrieval processes (Johnson et al., 2009). Despite these differences, the present experiment established similar ERP effects to Bridge and Paller (2012) and Liu et al. (2018) that corresponded to the accurate retrieval face memories. However, it should be emphasised that precise mechanisms that underpin the recognition-induced updating of face memories may still qualitatively differ to other forms of retrieval-induced updating. This point will be developed further in the general discussion (see Chapter 7).

A final discussion comment regarding the ERP findings from this chapter refers to the statistical power of the reported ERP effects. Like any ERP study, the signal-to-noise ratio of averaged ERPs is an important consideration (Luck, 2005) to ensure that averaged signals reflect reliable as opposed to random fluctuations in time-locked EEG activity. Many recommendations on suitable sample size, and number of trials per participant/condition, have been published to ensure an optimal signal-to-noise ratio of ERPs. Whilst traditional recommendations suggested a fixed number of trials per ERP condition (Luck, 2005), more recent suggestions imply that maximising signal-to-noise ratios is determined by a mixture of sample size, trial numbers per condition, as well as the size of ERP effect of interest (Boudewyn et al., 2018). In the paper from Boudewyn et al. (2018), the authors show that (with a sample size of 30, the same as in the current experiment) with 8 trials per condition, power values of 80% plus are found only for ERP difference effects of 4-7µV. In the present experiment, some ERP contrasts differed only by 1-2µV (including interaction contrasts, which is particularly problematic as the noise of a double subtraction, i.e. interaction effect, is doubled) possibly implying that some ERP effects in the present experiment may have been underpowered, particularly for individual participants who had ~10 trials per condition. Therefore, a future adaptation of the present experiment may consider increasing the number of trials per condition for eligible participants, to ensure that ERP effects are sufficiently powered to detect an effect if one is present.

The behavioural data from Experiment 5 can be compared to results seen in Chapter 3, particularly the results from Experiment 3b which was almost identical to Experiment 5, with the exception that trials were now grouped into five blocks of 15 face sets, repeated across two sessions (compared to two blocks of 30 face sets in Experiment 3b) and the face stimuli used. Furthermore, as there were fewer face targets to memorise per block in Experiment 5, a shortened learning task was adopted to that used in Experiment 3b (see Appendix A). The first replication of behavioural data compared recognition accuracy and confidence as a function of initial (i.e. Test 1), repeated (i.e. Test 2 repeated) and delayed recognition (i.e. Test 2 baseline). Firstly, recognition confidence appeared to be larger during Test 1 vs. Test 2 conditions in Experiment 5, however these differences were statistically unreliable (a similar outcome to results from Experiment 3b).

For recognition accuracy, performance was highest at Test 1 compared to both Test 2 conditions. However, accuracy was significantly larger for the repeated vs. baseline condition, a finding inconsistent with data from Experiments 3a and b (where accuracy was not different between these two conditions). One reason why retrieval practice benefits were observed in Experiment 5 but not 3a-b could be that targets in Experiment 5 were not encoded as strongly (due to fewer learning cycles, or due to properties of the different face sets), meaning that retrieval practice would be more beneficial in strengthening these target face representations (c.f. Schapiro et al., 2018). Though, this idea does not hold when considering data from Appendix A, as the recognition accuracy was similar for the learning cycles used in Experiment 5 and 3a-b. Furthermore, recognition accuracy in Test 1 of Experiment 5 was 52%, much higher than accuracy rates during Test 1 of Experiment 3a-b (39% & 37% respectively). Focusing on these patterns of Test 1 performance, it may be the case that a small retrieval practice benefit was observed in Experiment 5 due to this increased accuracy, as suggested in research from the retrieval practice literature where retrieval practice is thought to be more likely when accuracy during initial retrieval is high (Smith et al., 2013).

In addition, behavioural data from Experiment 5 assessed patterns of recognition bias during Test 2 as a function of accuracy (correct vs. incorrect) and confidence (high vs. low) of responses during Test 1. Different patterns of bias between high and low confidence conditions were found between Test 1 correct vs. incorrect conditions. For correct Test 1 responses, participants were just as likely to select targets again during Test 2 regardless of whether the response on Test 1 was made with high or low confidence. In contrast, high vs. low recognition confidence for distractors selected in Test 1 increased the likelihood that the same distractor would be recognised during Test 2. Overall, these results are partially consistent with those seen in Experiments 3a-b. The pattern of bias for correct Test 1 responses conflicts with data from Experiments 3a-b (Test 2 bias was increased for correct Test 1 responses made with high vs. low confidence). The result from Experiment 5 is unlikely to be due to the median split procedure to create high and low confidence conditions for Test 1 responses, as mean confidence was larger for trials in the correct high compared to correct low confidence conditions (see Appendix E.3). Instead, it is possible that repeatedly selecting targets during Test 2 was not influenced by the confidence of Test 1 responses because targets may have been more memorable in Experiment 5 (due to the shorter delay between learning and recognition tasks, and fewer targets to learn per block). Therefore, confidence may only boost repeated correct recognitions of faces at longer time delays between encoding and retrieval. In contrast, when recognition errors were made during Test 1, high rather than low confidence responses possibly facilitated the encoding of the distractors into memory, increasing the likelihood that these faces would be recognised again at subsequent retrieval.

In summary, the results from Experiment 5 demonstrated that ERP positivities were associated with retrieval of episodic face memories that were correctly remembered with high confidence across multiple retrieval attempts, and may therefore index face memory reactivation. However, no ERP markers were shown to correlate with the encoding of errors made during recognition that updated face memories. One reason for these null effects may be due to limitations of the ERP technique in analysing EEG data. ERPs represent the time-locked activity of evoked EEG responses to stimuli, such as faces. It is also thought that ERPs are predominantly generated by slow-frequency oscillations in the delta, theta and alpha band (Bastiaansen et al., 2011). Consequently, ERPs do not capture induced EEG activity, and activity in higher frequency bands. One approach that may be able to better detect the neural mechanisms of retrieval-induced updating is to analyse EEG data by decomposing it into time-frequency representation of oscillations in different frequency bands. Time-frequency analysis has several advantages over time-domain analysis of EEG data, such as ERPs (see Chapter 2, section 2.4.4 for detailed discussion). It is possible that analysing oscillatory activity in the EEG experiment from Chapter 5 may reveal neural mechanisms of face memory retrieval and updating that could not be captured by the ERP method, as addressed in the next Chapter 6.

Chapter 6 - Oscillatory correlates of objective, subjective and updating processes during face memory retrieval

In Chapter 5, it was shown that late, sustained ERP positivities were elicited during correct recognition of target face memories when those faces were recognised with high confidence in both tests. However, no ERP effects were found during Test 1 that related to updating of memories as measured by repeated responses in Test 2, thus no neural markers of retrieval-induced updating were found. However, the absence of any ERP correlates of updating does not rule out that neural mechanisms were active during Test 1 to update memories, as the ERPs only corresponds to the time-locked, evoked EEG signal that is mostly generated by slow-wave EEG frequencies (Bastiaansen et al., 2011). Thus, neural markers of retrieval-induced updating may be found in the non-evoked EEG, or EEG from higher frequencies. The goal of Chapter 6, therefore, was to re-analyse the EEG data from Experiment 5 with time-frequency decomposition into oscillatory measures, in order to further investigate the neural mechanisms underlying face memory retrieval and updating.

EEG oscillation analysis in the context of episodic memory research has expanded in recent decades. From this literature, multiple different oscillatory mechanisms have been proposed that correlate with successful episodic memory encoding and retrieval. Typically, oscillatory activity in the theta (4-8Hz) and gamma bands (30-100Hz) is shown to increase (i.e. a synchronisation response) during successful episodic memory encoding (Friese, et al., 2013; Khader et al., 2010; Hanslmayr et al., 2011; Meeuwissen et al., 2010; Osipova et al., 2006) and retrieval (Jacobs et al., 2006; Klimesch et al., 2000; Wilding & Ranganath, 2011). Furthermore, the theta and gamma oscillations are thought to couple during memory formation and retrieval, with theoretical accounts (Parish et al., 2018) suggesting that theta oscillations generated by the hippocampus are involved in binding (during encoding) and reactivating (during retrieval) mnemonic representations in neocortical regions (represented by gamma power) that are integral to successful memory processing (Kerren et al., 2018; Nyhus & Curran, 2010).

In contrast, EEG oscillations in the alpha (8-12Hz) and beta bands (12-30Hz) have been proposed to reflect an inhibitory mechanism in the brain. Alpha and beta power decreases (i.e. a "desynchronisation" response) have been frequently associated with successful episodic memory encoding (Hanslmayr et al., 2009; Hanslmayr et al., 2011) and retrieval (Khader et al., 2010; Meeuwissen et al., 2010; Poch et al., 2018; Samuel et al., 2018), particularly in regions of the brain that are specific to the content of episodic memories (Graetz et al., 2019; Guran et al., 2019; Hanslmayr et al., 2016; Khader et al., 2010; Parish et al., 2016; Waldhauser et al., 2012). The material-specific nature of alpha/beta desynchronisation suggests a 'release of inhibition' in these regions to facilitate memory encoding and retrieval. However, alpha power increases have also been related to successful memory formation, particular during periods where items are maintained in working memory (Khader et al., 2010; Meeuwissen et al., 2010; Poch et al., 2018; Samuel et al., 2018), to protect to-be-remembered information from interference during memory formation.

Lastly, the oscillatory correlates of episodic memory updating have also been examined in the context of a directed forgetting manipulation. Bäuml et al. (2008) observed increased alpha power at temporal electrodes during encoding a second list of words that predicted whether participants would remember these words at the expense of words learnt from a previous list. This alpha power enhancement was thought to reflect active inhibition of list 1 memories as a result of intentional forgetting instructions that facilitated encoding of memories in list 2. Consequently, alpha synchronisation may functionally relate to the updating of episodic memories by inhibiting previously learned memories during the encoding of novel episodes. However, the research by Bäuml et al. (2008) is, to my knowledge, the only study to examine the oscillatory correlates of episodic memory updating. Furthermore, this study focused on intentional forgetting and updating of semantically rich word stimuli with elaborate item-context associations, thus the oscillatory mechanisms of updating may differ for simpler forms of memory retrieval of memories that have minimal semantic/conceptual information, such as recognition of faces. The research presented in the current chapter applied time-frequency analysis to the EEG data collected in Experiment 5. The goal of this analysis was to examine the oscillatory correlates of objective retrieval success, subjective retrieval experience and updating-related processes during the repeated retrieval of face memories.

As a reminder, the repeated recognition paradigm used in Chapter 5 initially presented participants with several face targets to memorise during a learning task. Following a short filler task, participants completed several recognition trials in Test 1. For each trial, one target face was shown amongst four distractor faces from the corresponding face set to the target. Participants were asked to select the face they recognised from the learning phase and provided confidence judgements for each recognition response. The same recognition task was completed during test 2, for ‘repeated’ trials that were shown during Test 1, and baseline trials that were not shown during Test 1. With this paradigm, EEG oscillations in theta, alpha and beta bands during Test 1 and 2 were analysed using the same approach and experimental conditions (using the exact same trials) as the ERP analysis presented in Chapter 5. The first analysis compared EEG oscillations as a function of retrieval accuracy (correct target recognition versus incorrect distractor recognition) and recognition confidence (high versus low confidence). It was expected that theta synchronisation at frontal or parietal sites, and alpha/beta desynchronisation at posterior electrodes (overlapping visual cortex) would distinguish accurate from erroneous recognition (e.g. Hanslmayr et al., 2015; Nyhus & Curran, 2010; Waldhauser et al., 2012; Wilding & Ranganath, 2011). These oscillatory effects may also distinguish between high and low confidence judgements, particularly within in the theta band at parietal locations (e.g. Wynn et al., 2019). However, a more exploratory hypothesis examined an interaction between accuracy and confidence across each frequency band, to determine whether EEG oscillations of confidence applied to both accurate and erroneous recognition judgements, something not yet studied in the literature.

A second analysis analysed EEG oscillations during Test 1 and 2 according to the similarity of recognition responses made in the alternate test, separately for accurate and inaccurate trials. In Test 1, oscillatory effects contrasted recognition according to whether participants made a similar response in Test 2 (future close bias) or made a dissimilar response in Test 2 (future far bias). For Test 2 trials, oscillations contrasted activity when participants repeated a similar response to that in Test 1 (repetition close bias) or made a dissimilar response to that in Test 1 (repetition far bias). Because of the novelty of this investigation, it was less clear how EEG oscillations in different frequency bands might be modulated by these comparisons. Nevertheless, it was assumed that oscillatory correlates of future bias during Test 1 should reflect encoding-related activity, since this comparison should covary with more (close bias) or less (far bias) encoding of faces selected during Test 1 that biased subsequent retrieval attempts in Test 2. During Test 2, oscillatory correlates of repetition bias were expected to reflect the retrieval and reactivation of face memories that had been updated in Test 1. Reactivation of updated memories in Test 2 would be expected to occur whenever participants gave a similar response as in Test 1 (close bias) compared to when they gave a dissimilar response (far repetition bias). It may be expected that reactivation-related oscillatory correlates would be larger for correct responses in Test 2 (that were strengthened by retrieval in Test 1) compared to the reactivation of distractor face memories that were encoded during Test 1.

6.1. Method

The current Chapter 6 presents a reanalysis of EEG data collected in Experiment 5, using exactly the same EEG trials and conditions. The methodological details of Experiment 5, including participants, face stimuli, procedure and EEG recording and preprocessing can be found in the Chapter 5 Methods section (Section 5.1).

Data analysis

The present chapter focused on analysing the oscillatory EEG correlates of recognition accuracy, confidence and bias, during both Test 1 and Test 2. As a reminder, oscillatory data was analysed according to; a) Test 1 recognition accuracy and confidence, b) Test 2 accuracy and confidence, c) Test 1 accuracy and future bias and d) Test 2 accuracy and repetition bias (see Chapter 5 for full description of how conditions were formed). Between test analysis were also conducted for the accuracy and confidence conditions (reported in Section 6.2) and accuracy and bias conditions (see Appendix F). Each of these four analyses were performed separately for three frequency bands of interest; theta (4-7Hz), alpha (8-12Hz) and beta (13-30Hz). As mentioned in Chapter 2, gamma frequencies were not analysed as frequencies above 40Hz were filtered out of recorded EEG.

Time-frequency decomposition. All time-frequency analysis was conducted using the Matlab toolbox FieldTrip (Oostenveld, Fries, Maris & Schoffelen, 2011), as well as custom-Matlab scripts. For each condition, epochs were decomposed into a time-frequency spectra using Morlet wavelets with a width of 3 cycles. Decompostition was performed between 1-30Hz, in steps of 1Hz, beginning from -1200ms pre-stimulus to 1600ms post-stimulus. As the task design necessitated relatively short duration epochs (-0.4 to 1.6s), an initial step before decomposition was to extend the pre-stimulus baseline period from -0.4s to -1.2s of the raw EEG epochs, in order to accurately measure slower-frequencies of the EEG time-frequency spectrum. The 'flipud' Matlab function was applied to the raw EEG signal recorded during the -0.4-0s pre-stimulus time window for each epoch in order to invert this time window, which was then appended to the start of the epoch. The original baseline time window was then appended to the start of the flipped time window, resulting in pre-stimulus time periods of -1.2s-0s. The purpose of including the inverted signal in the middle of the new pre-stimulus period was to prevent sudden "jumps" in amplitudes where segments were appended (see Figure 2.3 in Section 2.4.4), which could otherwise cause artefacts in the EEG decomposition (following previous research, e.g. Vogelsang, Gruber, Bergström, Ranganath & Simons, 2018; see Cohen, 2014). These extended epochs (-1.2s – 1.6s) were then entered into time-frequency decomposition, in order to estimate power of both evoked and induced oscillations. EEG data was normalised to a dB scale and baseline corrected to the -825ms to -375ms pre-stimulus time period[1]. Oscillatory power at the different frequencies was then averaged across trials in the experimental conditions at the individual level for 30 participants eligible for analysis, and these averages were used for statistical analysis.

Cluster-based permutation statistical analysis. For each of the four main analyses, cluster-based permutation tests were performed with mean dB scaled power as a

---

[1] The baseline times were selected to truncate 375ms from the start and end of the pre-stimulus baseline because the temporal width of a 3 cycle wavelet at the lowest frequency (4Hz) is 750ms. Therefore, -825ms was the last time point where the wavelet width (at 4Hz) would include all time points (e.g. at -826ms and higher, the left tail of the wavelet would be a time point of 1201ms which was beyond the epoch limit). Similarly, -375ms was the earliest time point where the wavelet width (at 4Hz) would include time points solely from the baseline time window (e.g. at -324ms and lower, the right tail of the wavelet would be a time point of 1ms which would include early activity related to the perception of face stimuli; see Section 2.4.4 in Chapter 2).

dependent variable, collapsed across frequency values within respective frequency bands (theta - 4-7Hz, alpha - 8-12Hz, beta - 13-30Hz) in order to compare the effects of accuracy, confidence and bias between discrete frequency band boundaries (rather than across a continuous frequency spectrum) . Within each of three analysed frequency bands, paired samples t-tests (two tails, alpha threshold = .05) were performed between two conditions at 60 electrodes (VEOG and HEOG excluded) from 0-1.1s[2], for main effects, interaction and simple main effect comparisons. For the first step of the cluster tests, a t-test value for spatio-temporal data samples lower than the critical alpha threshold (0.05) was included in a cluster, so long as the t-test for at least 2 neighbouring spatio-temporal samples was also lower than the alpha threshold. In the second step, cluster-level statistics were calculating by summing t-values for data samples included in the cluster. The size of clusters then underwent significance testing by creating a permutation null distribution (5000 resamples). The proportion of permuted t-values larger than the observed cluster-level statistic in this distribution provided the significance (*p*-value) of the cluster-level statistics (two-tails, cluster alpha threshold = .05). Clusters that had a *p*-value below the cluster alpha threshold were interpreted on spatial and temporal dimensions using line and topographical plots.

6.2. Results

Omnibus analysis and simple main effects analysis of accuracy and confidence, as well as accuracy and bias, were performed separately for each frequency band of interest (theta, alpha & beta) separately for Tests 1 and 2. In addition, conditions were compared

---

[2] As in the corresponding ERP analysis in Chapter 5, the analysed time window was shorter than the total face presentation time as errors in the PsychoPy scripts meant that a small number of face presentations were shortened by the next fixation cross being presented earlier than expected. Thus, time windows of 1.1s were analysed to avoid the risk of analysing ERP activity that did not correspond to faces on-screen.

between tests to assess whether reliable differences of oscillatory power for accuracy, confidence and bias contrasts emerged as a function of repeated retrieval.

Oscillatory correlates of accuracy and confidence

Test 1. Firstly, cluster-based permutation tests analysed theta, alpha and beta power differences for main effects of recognition accuracy, confidence, and an accuracy-confidence interaction for Test 1 (see Table 6.1 for cluster statistics). Here, three significant negative clusters were found in the beta band. Figure 6.1 depicts the largest cluster, showing differences between correct and incorrect conditions at approximately 1-1.1s, prominent at centro-posterior electrodes. The other two clusters suggest differences between correct and incorrect conditions at centro-posterior sites from ~0.8-0.9s (cluster two) and ~0.7s-0.8s (cluster three). For all three of these clusters, Figures 6.4 and 6.5 show that beta power was reduced for correct compared to incorrect conditions. Similarly, negative clusters in the alpha and beta band were shown to be significant for the main effect of confidence. Effects in both bands corresponded to a cluster where differences between high-low confidence conditions emerged at centro-posterior sites from approximately 0.8/0.9-1.1s. In both bands (see Figures 6.3 and 6.4), alpha/beta power was reduced for high confidence compared to low confidence conditions. Finally, a positive cluster for the interaction term was significant in the beta band, though this cluster lasted for less than 100ms from 0.7-0.8s thus no cluster points are plotted in Figure 6.1. A follow-up cluster test established no significant positive clusters for the simple main effect of correct high-correct low confidence contrast (cluster $t = 31.99$, $p = .88$), with no clusters formed for the incorrect high-incorrect low confidence contrast (which may have contributed to the positive cluster for the interaction being significant). Instead, negative clusters in the beta band were significant for both the simple main effects, as expected given that the largest negative cluster for the interaction effect was not significant (cluster *t* = -131.65, *p* = .52). Similar to the confidence main effect, clusters for correct high and incorrect high conditions showed reduced beta power, later in the time window predominantly at centro-posterior electrodes, compared to correct low and incorrect low confidence conditions respectively (see Figure 6.4), however this modulation by confidence was stronger for the incorrect conditions. Thus, these results show that late beta desynchronisation effects in Test 1 were found for high versus low confidence responses during both correct and incorrect recognition.

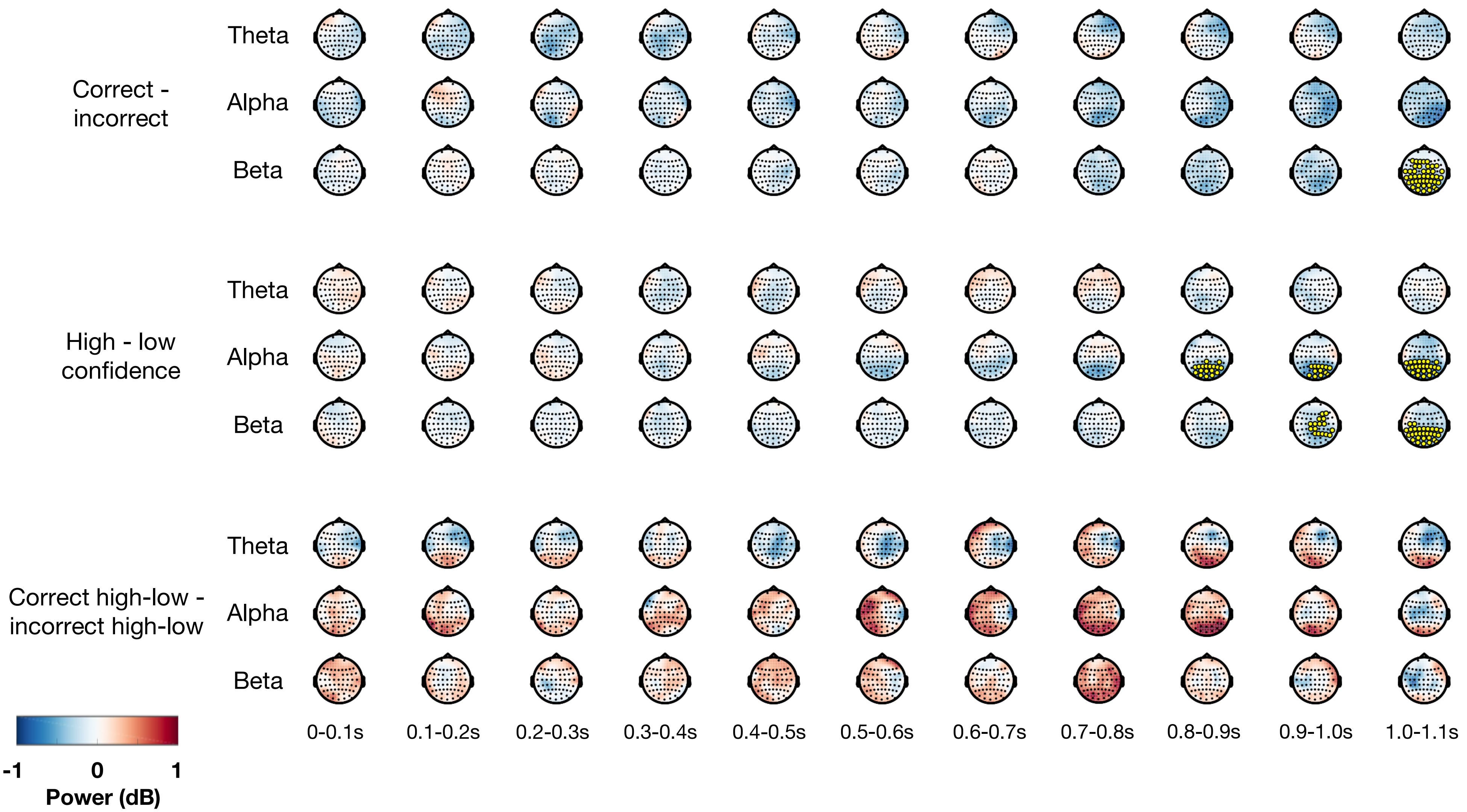

Figure 6.1. Cluster plots for oscillatory analysis of test 1 accuracy and confidence. Topographical maps with significant clusters (denoted by yellow dots) demonstrating main effects of accuracy and confidence, as well as an accuracy-confidence interaction in test 1 for theta, alpha and beta bandwidths. Only the largest clusters are plotted.

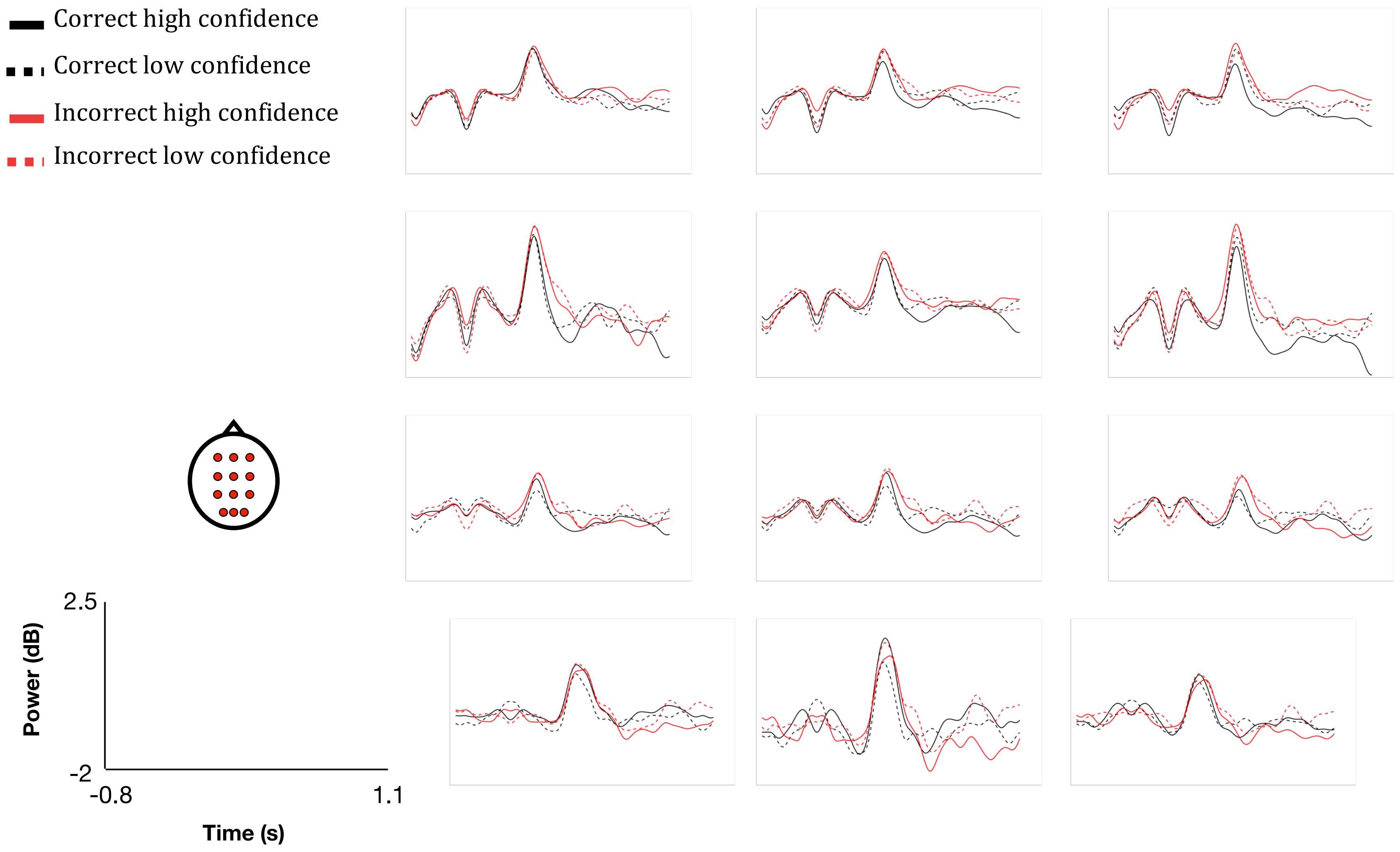

Figure 6.2. Time-frequency plots of theta power for test 1 accuracy and confidence. Line plots of theta power for correct high, correct low, incorrect high and incorrect low confidence conditions in test 1. Theta power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2.

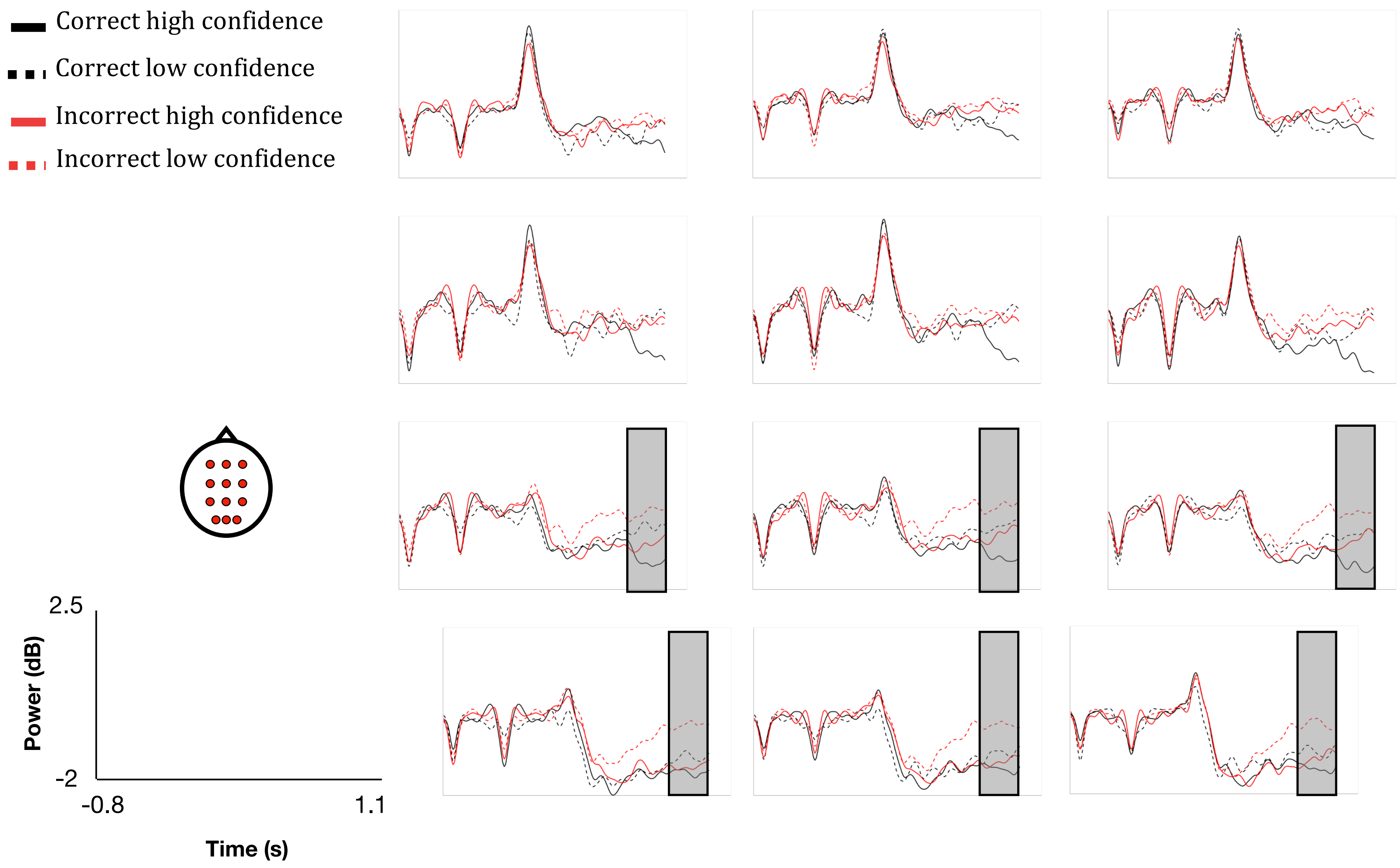

Figure 6.3. Time-frequency plots of alpha power for test 1 accuracy and confidence. Line plots of alpha power for correct high, correct low, incorrect high and incorrect low confidence conditions in test 1. Alpha power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections represent significant time points.

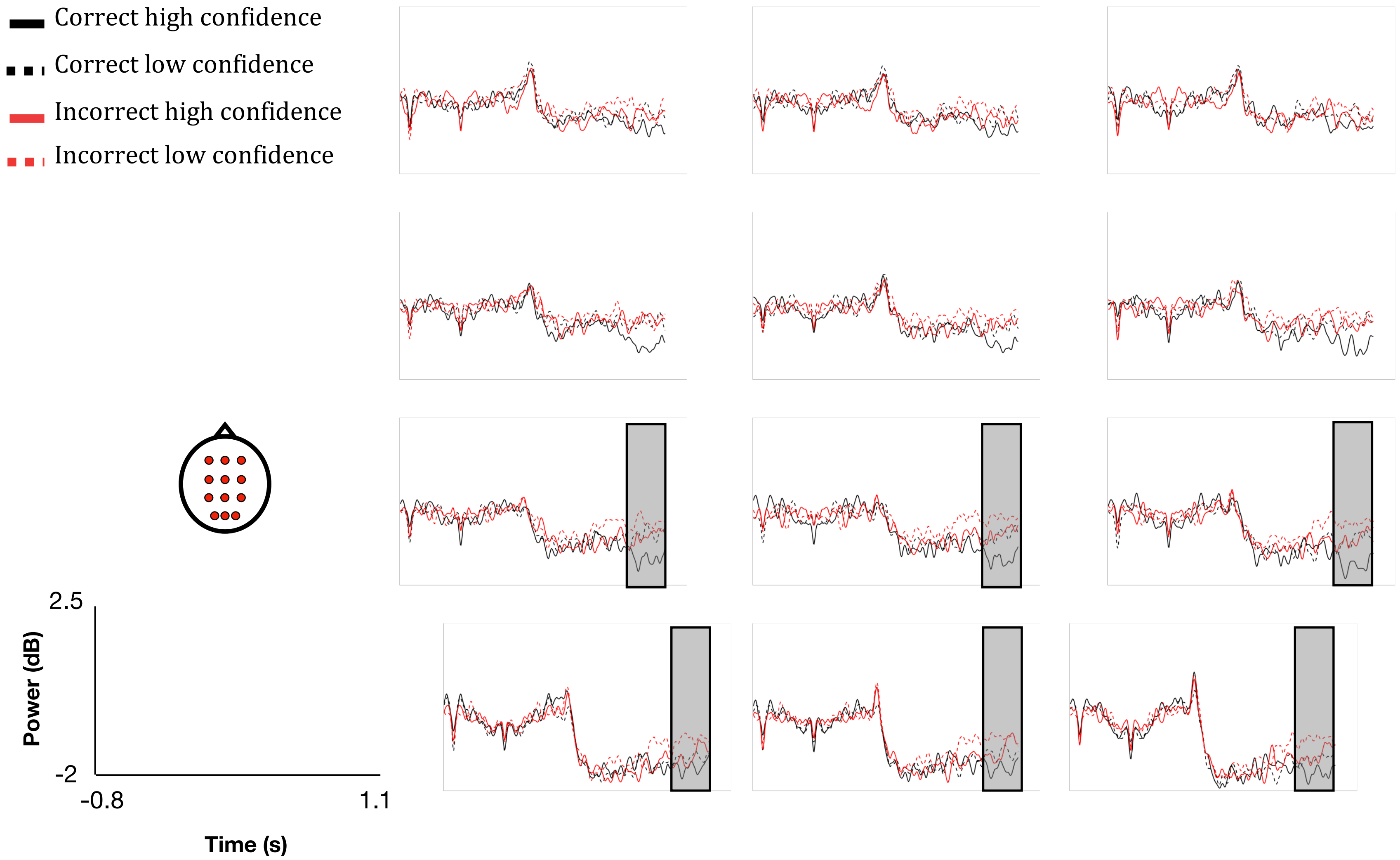

Figure 6.4. Time-frequency plots of beta power for test 1 accuracy and confidence. Line plots of beta power for correct high, correct low, incorrect high and incorrect low confidence conditions in test 1. Beta power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections represent significant time points.

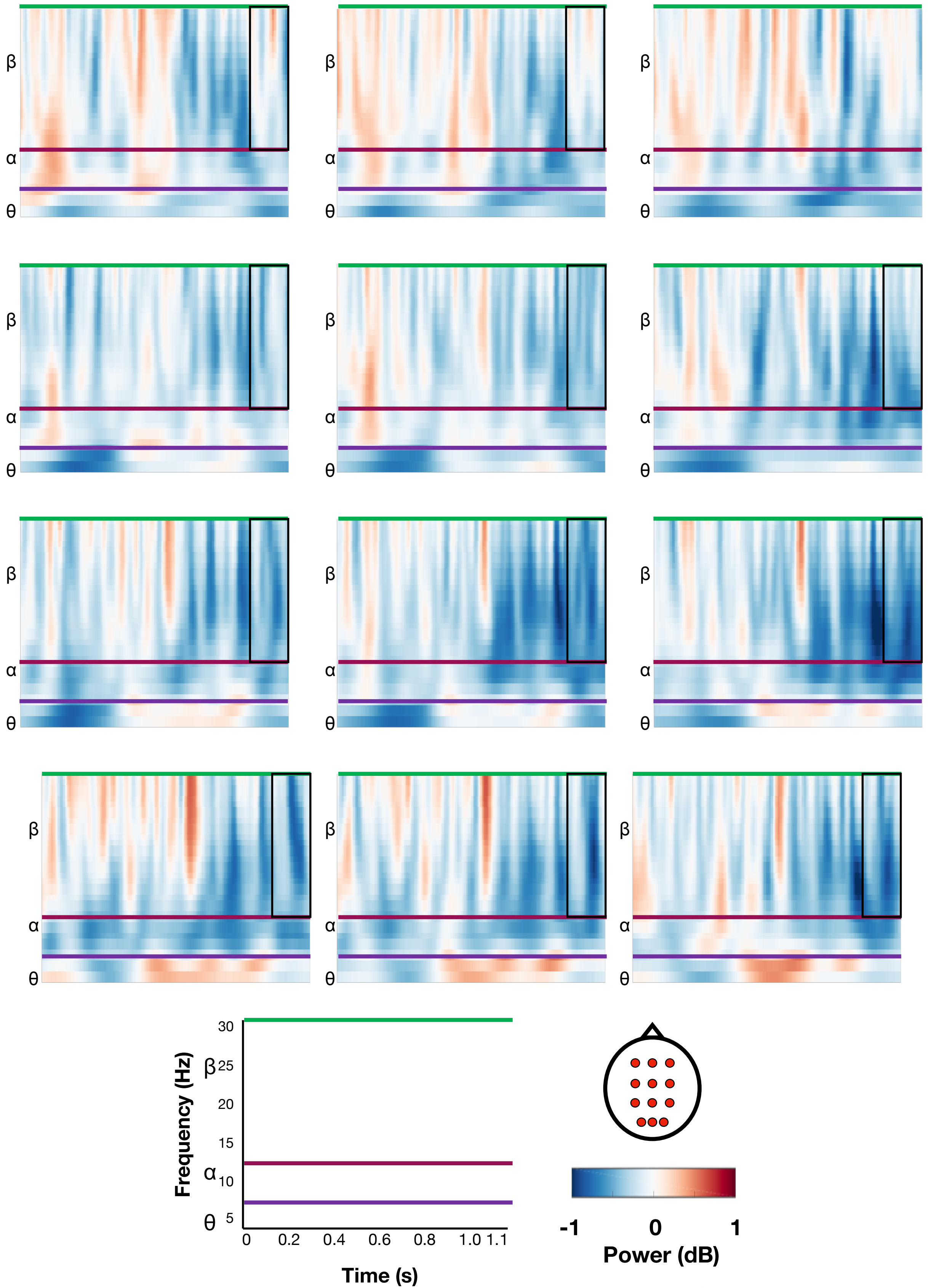

Figure 6.5. Time-Frequency representations for Test 1 accuracy. Plots denote power differences between correct and incorrect conditions in Test 1, plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections denote time, electrode and frequency data part of a significant cluster.

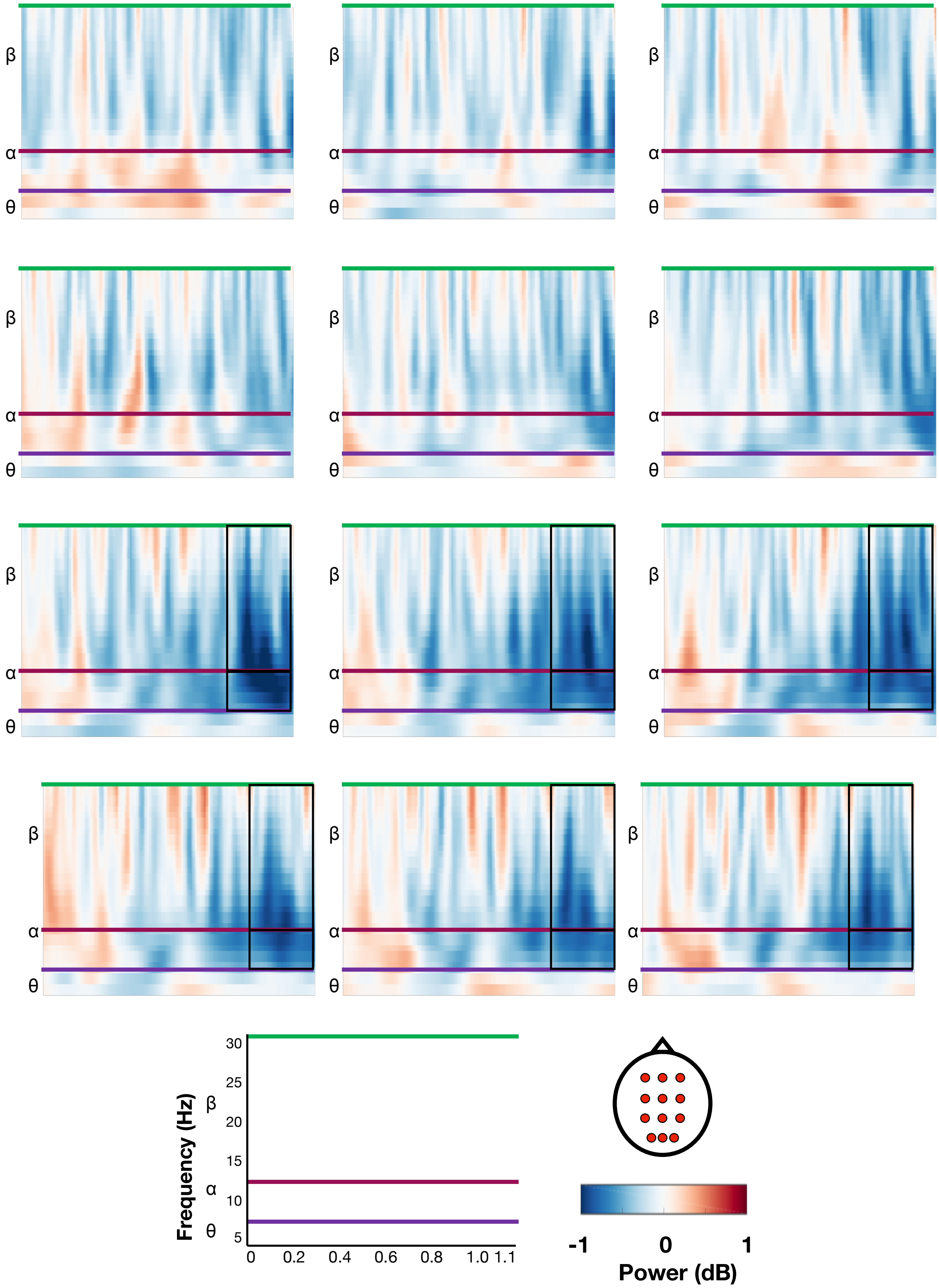

Figure 6.6. Time-Frequency representations for Test 1 confidence. Plots denote power differences between high and low confidence conditions in Test 1, plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections denote time, electrode and frequency data part of a significant cluster.

Table 6.1. Cluster statistics for main effects and interactions of Test 1 and Test 2 accuracy and confidence conditions for theta, alpha and beta bands. All significant clusters are reported. For effects with no significant clusters, the largest cluster (across positive and negative clusters) for each contrast is reported. Contrasts with no clusters formed are denoted as -.

	Test 1						Test 2					
	Theta		Alpha		Beta		Theta		Alpha		Beta	
Cluster contrast	Cluster t	*p*	Cluster t	*p*	Cluster t	*p*	Cluster t	*p*	Cluster t	*p*	Cluster t	*p*
Correct-incorrect	-965.78	.13	-1323.68	.07	-2123.76 -780.96 -692.31	.004 .04 .05	1598.10	.06	-459.30	.25	-484.11	.09
High-low confidence	-	-	-4145.55	.003	-2990.41	<.001	521.48	.22	-2430.95	.03	-2340.79	.002
Interaction	48.25	.53	444.96	.24	666.09	.04	-	-	-27.95	.62	-152.26	.42
Correct high-correct low	-	-	-918.84	.10	-972.95 -918.55	.01 .02	469.31	.25	-1293.05	.07	-2889.76	<.001
Incorrect high-incorrect low	-861.80	.15	-4359.03	<.001	-1219.58 -750.75 -741.93	.003 .02 .02	-	-	100.70	.48	-74.35	.62

Test 2. In Test 2, clusters for the main effect of recognition accuracy were not significant for any frequency band. In contrast, for the main effect of confidence, clusters in the alpha and beta band were significant (see Figure 6.7). Both clusters corresponded to a reduction in alpha/beta power for high confidence versus low confidence conditions from approximately 0.8-1.0/1.1s. As seen in Figures 6.9, 6.10 and 6.12, these effects were apparent across the whole scalp. Finally, the interaction effect showed no significant clusters in any frequency band. In summary, the analysis of Test 2 accuracy and confidence conditions largely revealed effects related to subjective retrieval processes, with alpha/beta power decreases related to high confidence during recognition.

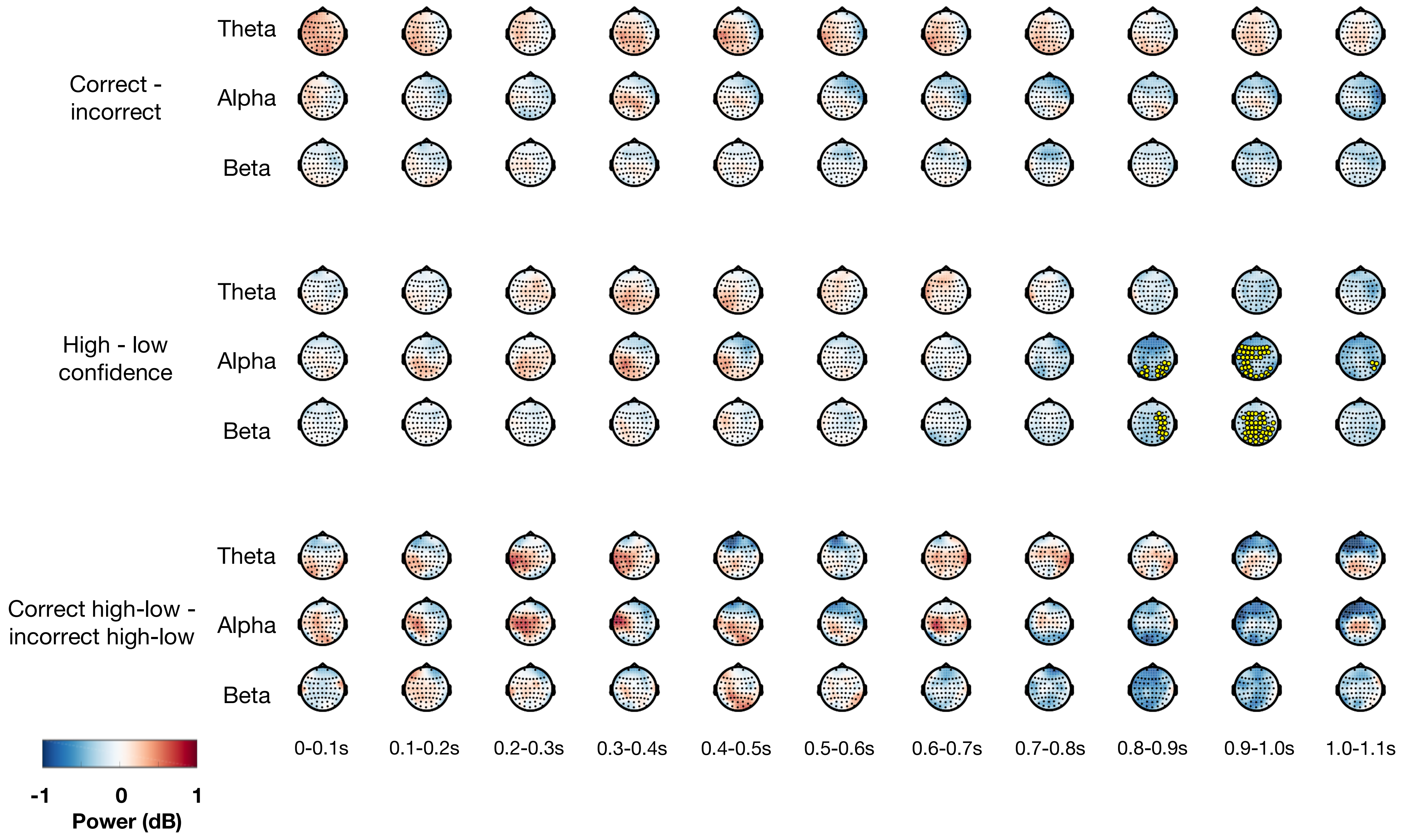

Figure 6.7. Cluster plots for oscillatory analysis of test 2 accuracy and confidence. Topographical maps with significant clusters (denoted by yellow dots) demonstrating main effects of accuracy and confidence, as well as an accuracy-confidence interaction in test 2 for theta, alpha and beta bandwidths. Only the largest clusters are plotted.

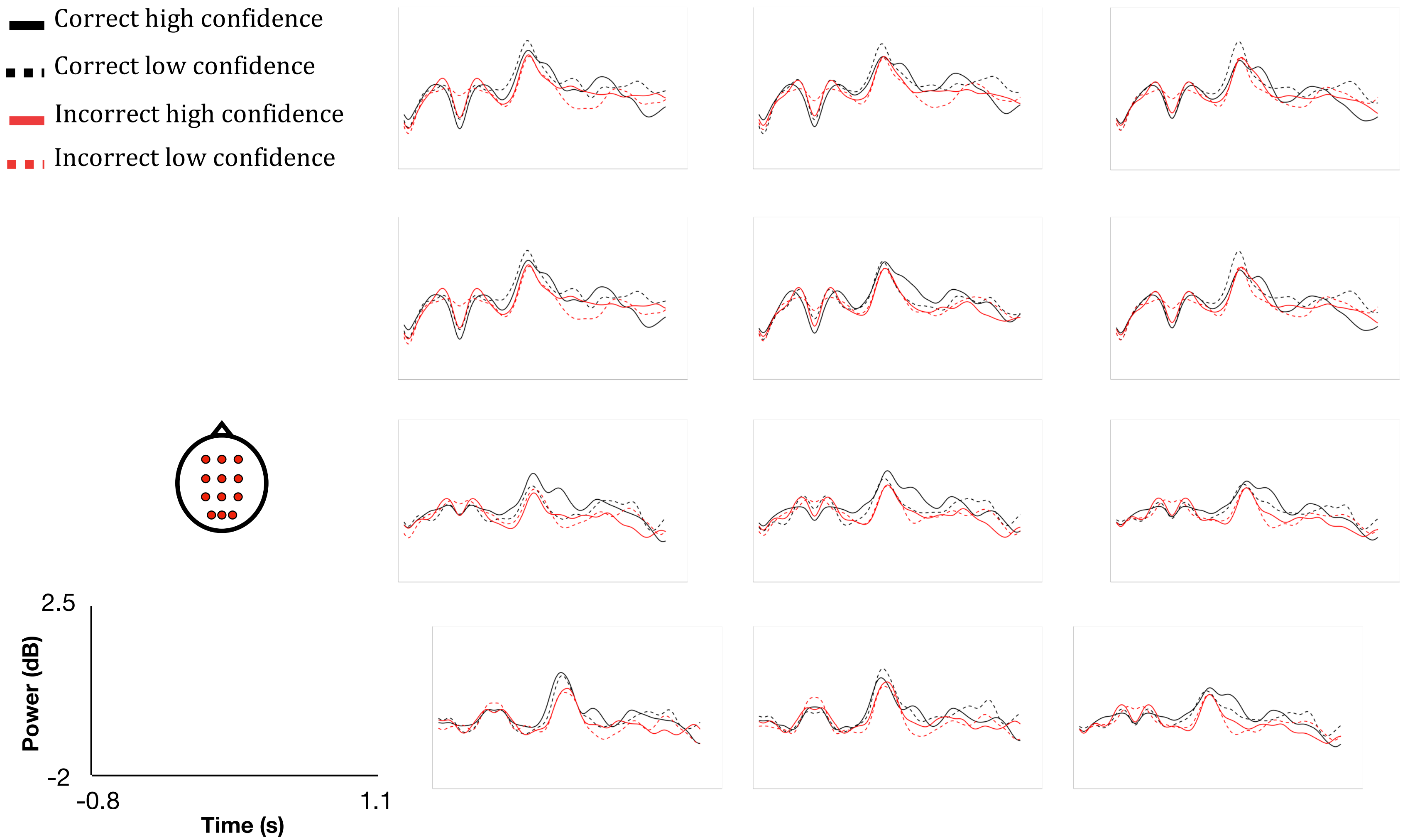

Figure 6.8. Time-frequency plots of theta power for test 2 accuracy and confidence. Line plots of theta power for correct high, correct low, incorrect high and incorrect low confidence conditions in test 2. Theta power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2.

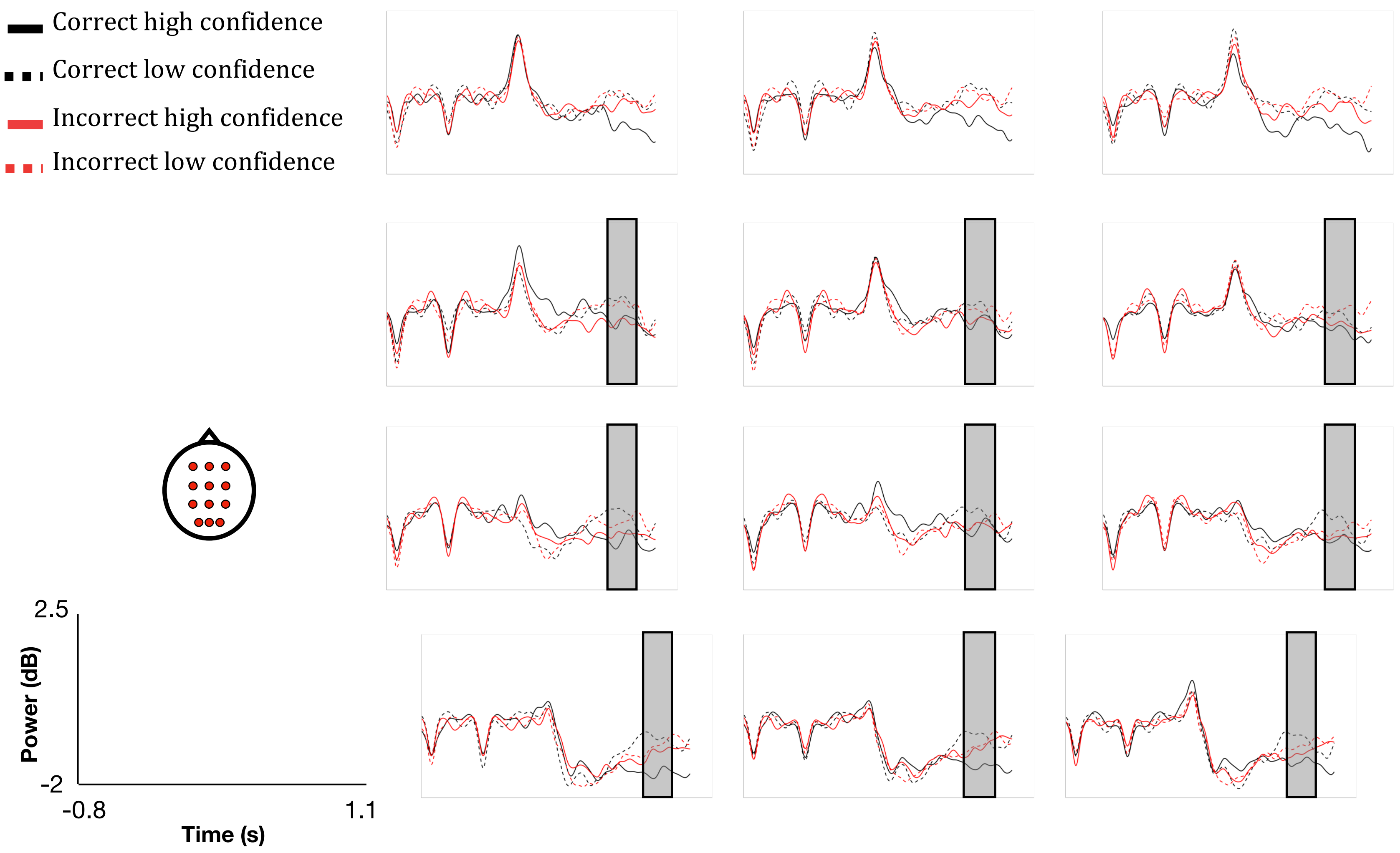

Figure 6.9. Time-frequency plots of alpha power for test 2 accuracy and confidence. Line plots of alpha power for correct high, correct low, incorrect high and incorrect low confidence conditions in test 2. Alpha power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections represent significant time points.

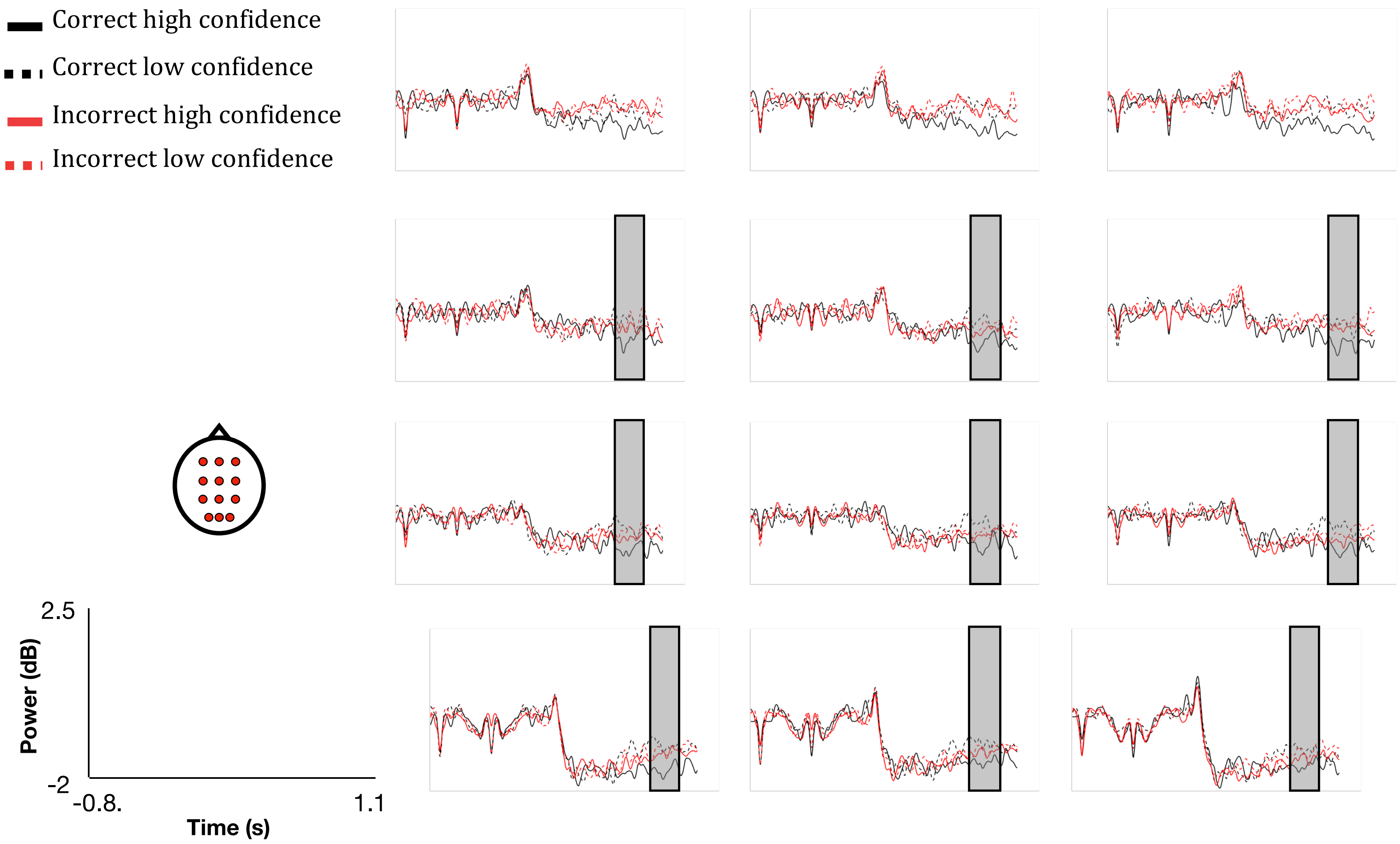

Figure 6.10. Time-frequency plots of beta power for test 1 accuracy and confidence. Line plots of alpha power for correct high, correct low, incorrect high and incorrect low confidence conditions in test 2. Beta power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections represent significant time points.

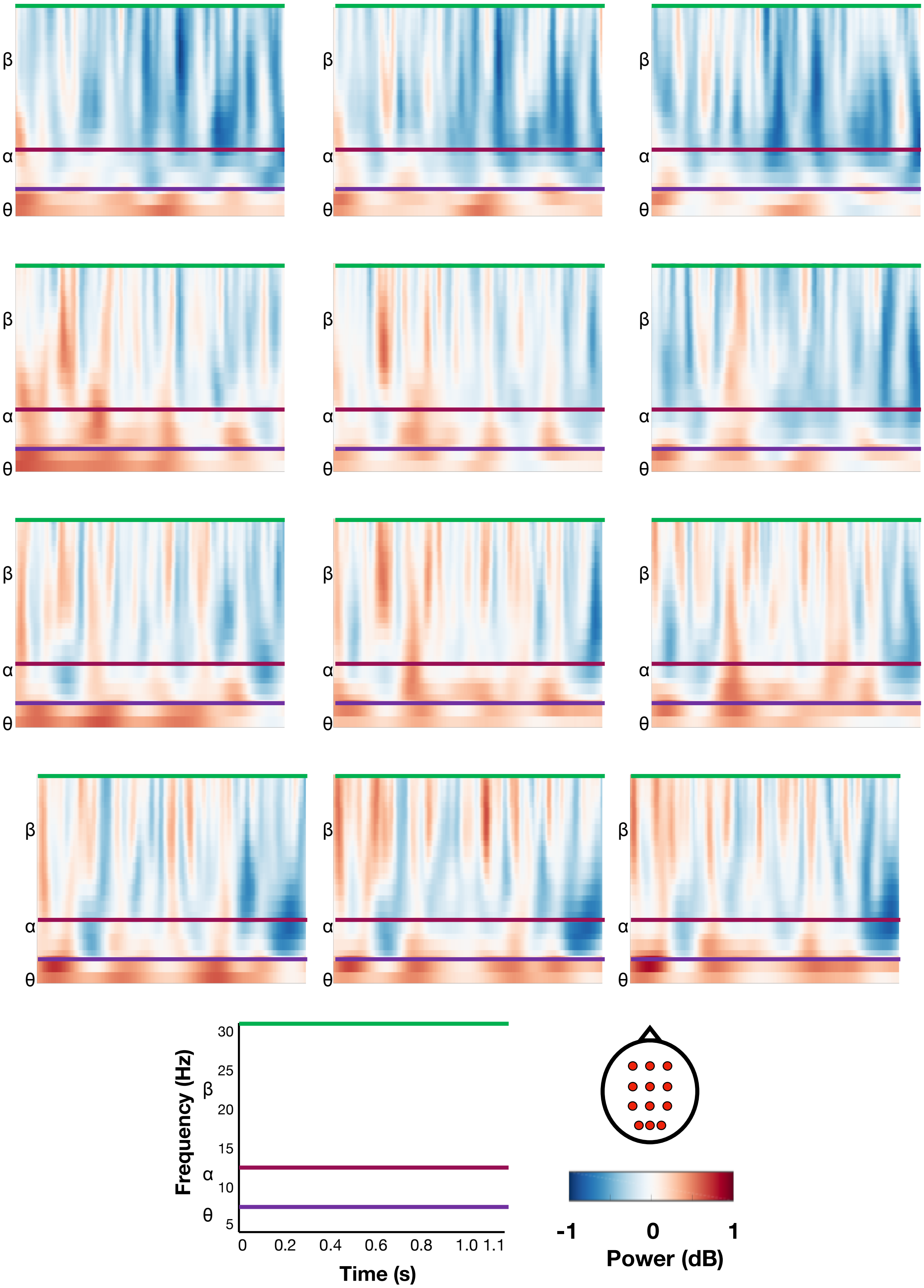

Figure 6.11. Time-Frequency representations for Test 2 accuracy. Plots denote power differences between correct and incorrect conditions in Test 2, plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2.

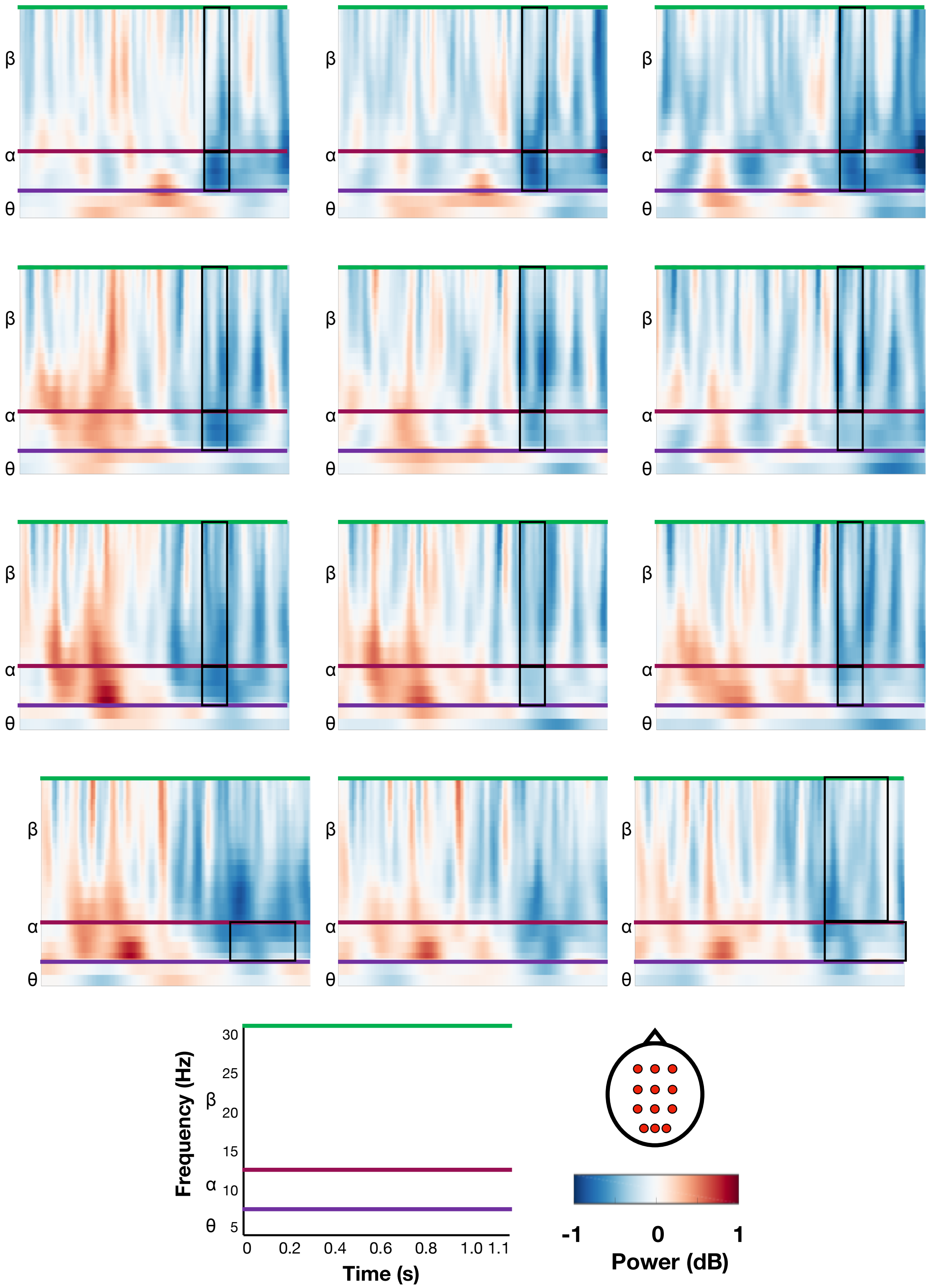

Figure 6.12. Time-Frequency representations for Test 2 confidence. Plots denote power differences between high and low confidence conditions in Test 2, plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections denote time, electrode and frequency data part of a significant cluster.

Test 1 vs Test 2. Finally, oscillatory analyses compared the theta, alpha and beta power for accuracy and confidence conditions between Tests 1 and 2. In this analyses, for each frequency band, a three-way design was adopted with factors of test (Test 1, Test 2), accuracy (correct, incorrect) and confidence (high, low). For this analysis, cluster corrected paired samples t-tests were performed between Tests 1 and 2 (see Table 6.2 for cluster test results), with dependent variables of the accuracy-confidence interaction (i.e. correct high-correct low confidence difference minus incorrect high-low confidence difference) as a three-way interaction term. Two-way interactions also compared the difference between correct and incorrect conditions, and high confidence and low confidence conditions, between tests. Finally, oscillatory power for the correct high, correct low, incorrect high and incorrect low confidence conditions were compared between Test 1 and 2.

Within the theta band, a significant cluster was found for the accuracy x test interaction. This negative cluster, corresponding to a centro-posterior effect from approximately 0-0.4s shows that the difference between correct and incorrect theta power was different in Test 2 compared to Test 1. As can be seen by comparing Figure 6.5 with Figure 6.11, early in the time window correctly recognised target faces were associated with increased theta power compared to incorrectly selected distractor faces in the second test, but not in the first test, where the theta power difference was reversed (lower power for correct than incorrect selections). In addition, a significant cluster was found for the three-way interaction in the beta band. This cluster, with a centro-posterior topography that lasted from approximately 0.6 to 0.8 seconds, shows that the difference between correct high-correct low confidence and incorrect high-incorrect low confidence conditions was larger in Test 1 (where there was a significant two-way interaction between accuracy and confidence for beta power) compared to Test 2 (where there was no such interaction).

Table 6.2. Comparison of time-frequency power of accuracy and confidence between Tests 1 and 2. Cluster corrected paired samples t-tests for the 3-way interaction between test, accuracy and confidence conditions. All significant clusters are reported. For effects with no significant clusters, the largest cluster (across positive and negative clusters) for each contrast is reported. Contrasts with no clusters formed are denoted as -.

	Theta		Alpha		Beta	
Cluster contrast	Cluster *t*	*p*	Cluster *t*	*p*	Cluster *t*	*p*
Accuracy x Confidence x Test	-	-	850.2	.12	1036.70	.007
Accuracy x Test	-5497.10	.007	-469.07	.21	-471.43	.11
Confidence x Test	-875.70	.14	-695.76	.15	190.89	.36
Correct high-low difference	-399.20	.24	-92.84	.51	542.65	.08
Incorrect high-low difference	7.26	.58	-886.19	.11	-474.96	.08

Oscillatory correlates of accuracy and bias

Next, the oscillatory correlates of recognition accuracy and bias were analysed in both Tests 1 and 2 to assess main effects of bias (close versus far bias) and an accuracy by bias interaction (see Table 6.3 for cluster statistics; note that the main effects of accuracy were already reported in the previous section so are included for comparison but not discussed in the text). As a reminder, in Test 1, oscillations elicited by selected faces were compared for trials where a similar face would later be selected in Test 2 (close future bias) and trials where a dissimilar face was selected in Test 2 (far future bias). During Test 2, oscillations were compared for trials when participants selected a face that was similar to their previously selected face in Test 1 (close repetition bias) versus trials when they selected a face that was dissimilar to the face they had selected in Test 1 (far repetition bias). During Test 1, it was predicted that oscillatory activity should differentiate close versus far bias conditions to reflect updating-related mechanisms, such as encoding of selected faces that predicted making a similar recognition response during Test 2. In Test 2, oscillatory differences were expected for the close and future bias conditions for both correct and incorrect Test 2 responses, providing evidence of the retrieval of updated face memories.

For Test 1, clusters for the main effect of future bias were significant in the alpha band (see Table 6.3). This positive cluster corresponded to increase alpha power for the close future bias condition from approximately 0.7-0.9s at left centro-posterior electrodes (see Figure 6.13). Furthermore, for the interaction term, significant positive clusters were found in alpha and beta frequencies. Follow-up simple main effects were conducted, with clusters in the alpha/beta bands showing increased power for future close versus future bias conditions for correct trials only. Clusters emerged from 0.7-1.0s primarily at centro-posterior electrodes for both alpha and beta frequencies (see Figures 6.15 & 6.16).

Table 6.3. Cluster statistics for main effects and interactions of accuracy and bias in Test 1 and 2 for theta, alpha and beta bands. All significant clusters are reported. For effects with no significant clusters, the largest cluster (across positive and negative clusters) for each contrast is reported. Contrasts with no clusters formed are denoted as -.

	Test 1						Test 2					
	Theta		Alpha		Beta		Theta		Alpha		Beta	
Cluster contrast	Cluster *t*	*p*	Cluster *t*	*p*	Cluster *t*	*p*	Cluster *t*	*p*	Cluster *t*	*p*	Cluster *t*	*p*
Correct-incorrect[1]	-965.78	.13	-1323.68	.07	-2123.76 -780.96 -692.31	.004 .04 .05	1598.10	.06	-459.30	.25	-484.11	.09
Close-far bias	1001.04	.12	1615.91	.05	439.09	.12	3334.45	.02	2238.47	.03	182.20	.36
Interaction	866.75	.15	2560.42	.02	1750.14	.004	-68.33	.46	690.26	.15	129.38	.43
Correct close-correct far	1893.16	.05	3171.54	.02	816.01	.04	1008.76	.12	2243.40	.03	192.20	.31
Incorrect close-incorrect far	356.99	.33	-396.47	.29	-322.40	.19	1033.78	.12	334.34	.34	67.19	.60

[1] The correct-incorrect contrast is identical to that seen in Table 6.1 because the same trials are allocated to the correct and incorrect conditions for both analyses.

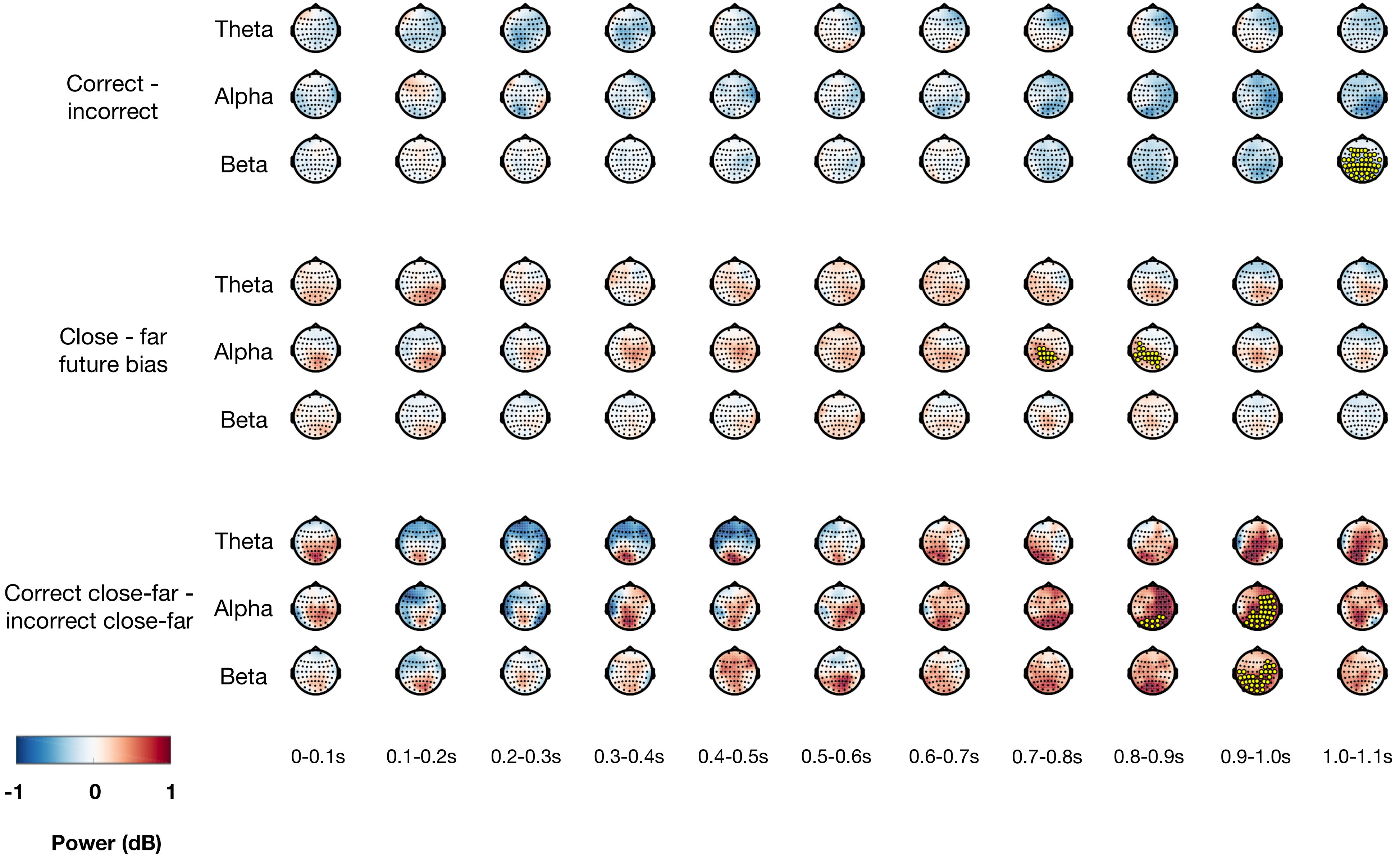

Figure 6.13. Cluster plots for oscillatory analysis of test 1 accuracy and future bias. Topographical maps with significant clusters (denoted by yellow dots) demonstrating main effects of accuracy and bias, as well as an accuracy-bias interaction in test 1 for theta, alpha and beta bandwidths. Only the largest clusters are plotted.

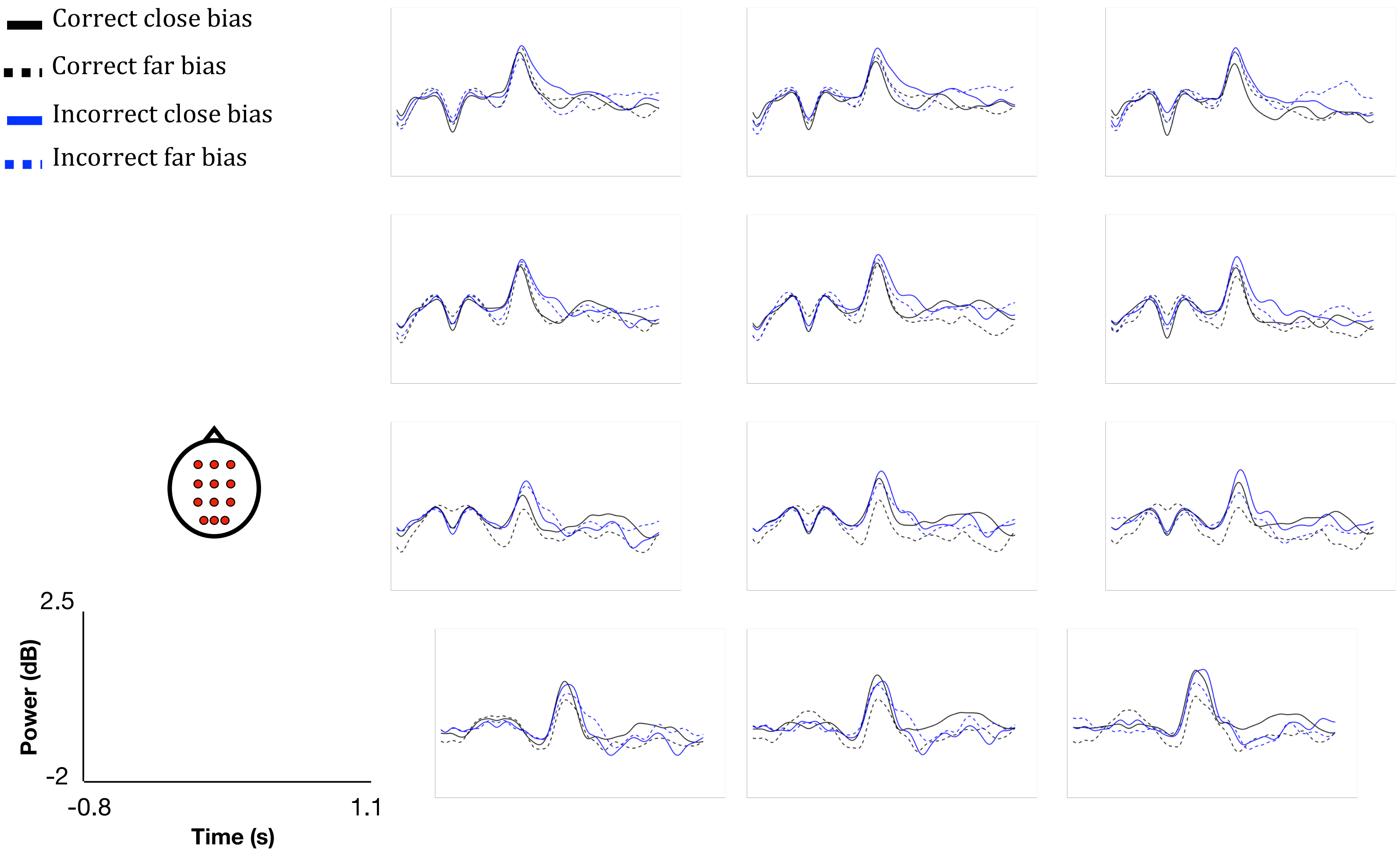

Figure 6.14. Time-frequency plots of theta power for test 1 accuracy and bias. Line plots of theta power for correct close, correct far, incorrect close and incorrect far bias conditions in test 1. Theta power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2.

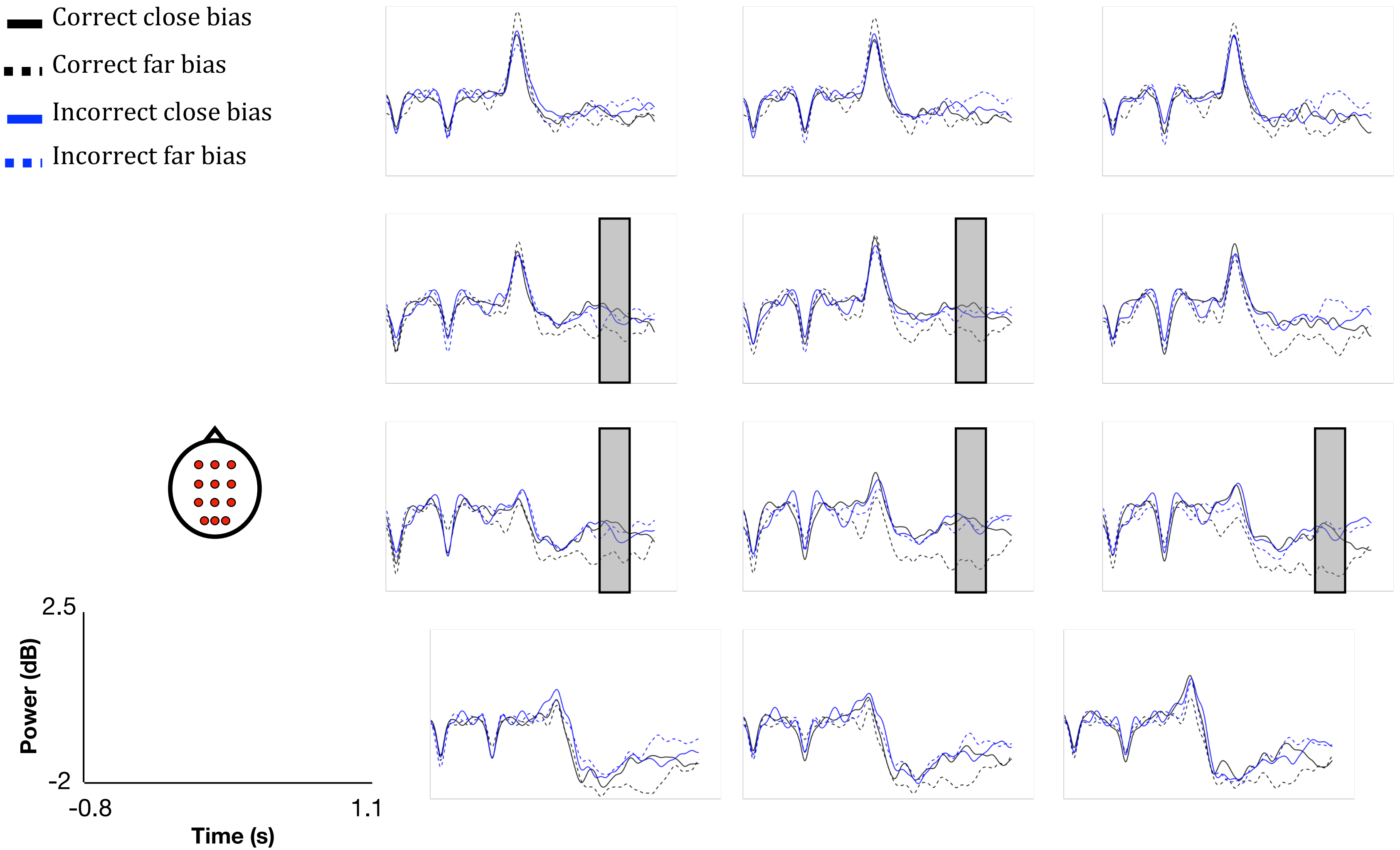

Figure 6.15. Time-frequency plots of alpha power for test 1 accuracy and bias. Line plots of theta power for correct close, correct far, incorrect close and incorrect far bias conditions in test 1. Alpha power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2 Shaded sections represent significant time points.

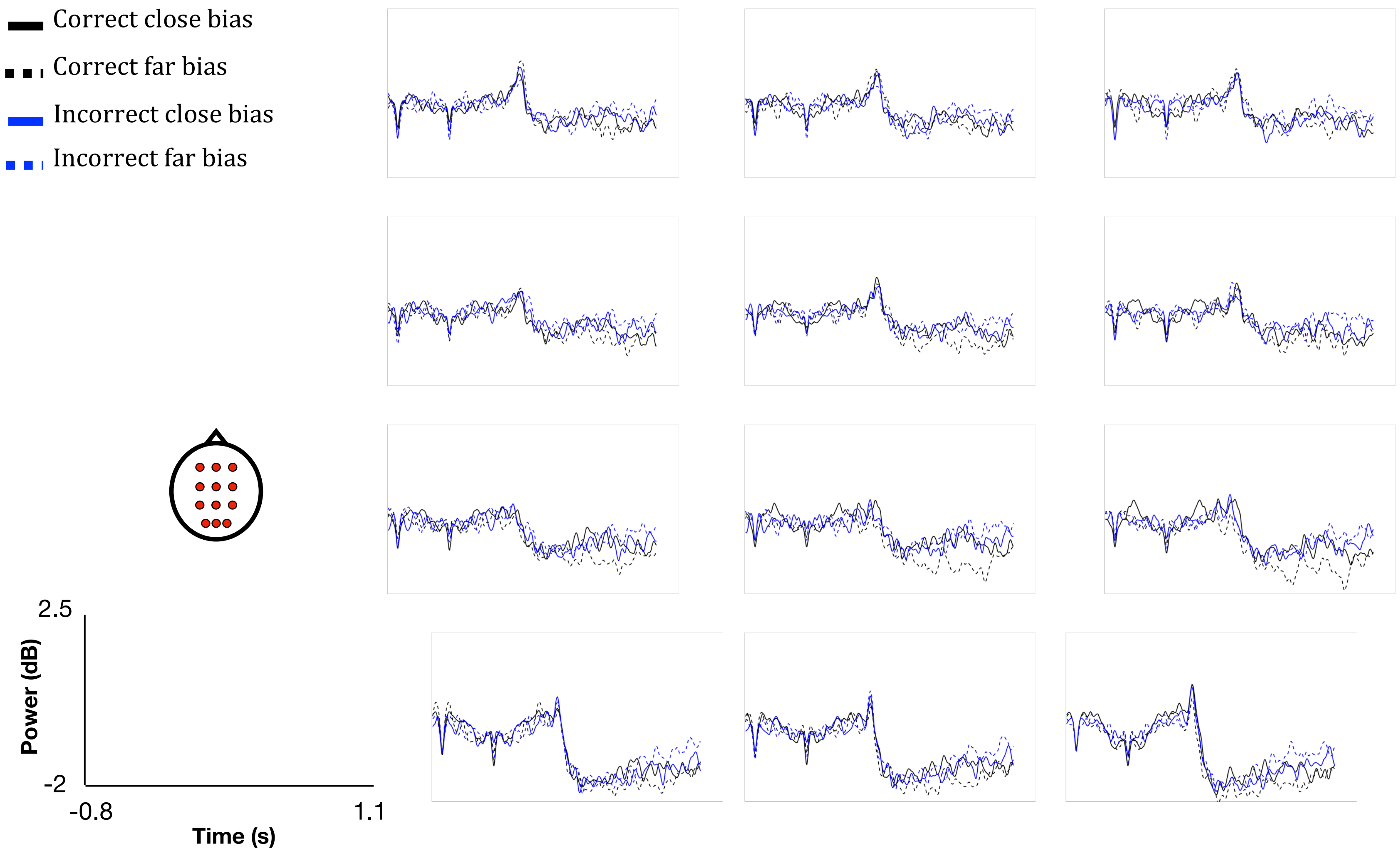

Figure 6.16. Time-frequency plots of beta power for test 1 accuracy and bias. Line plots of theta power for correct close, correct far, incorrect close and incorrect far bias conditions in test 1. Beta power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2.

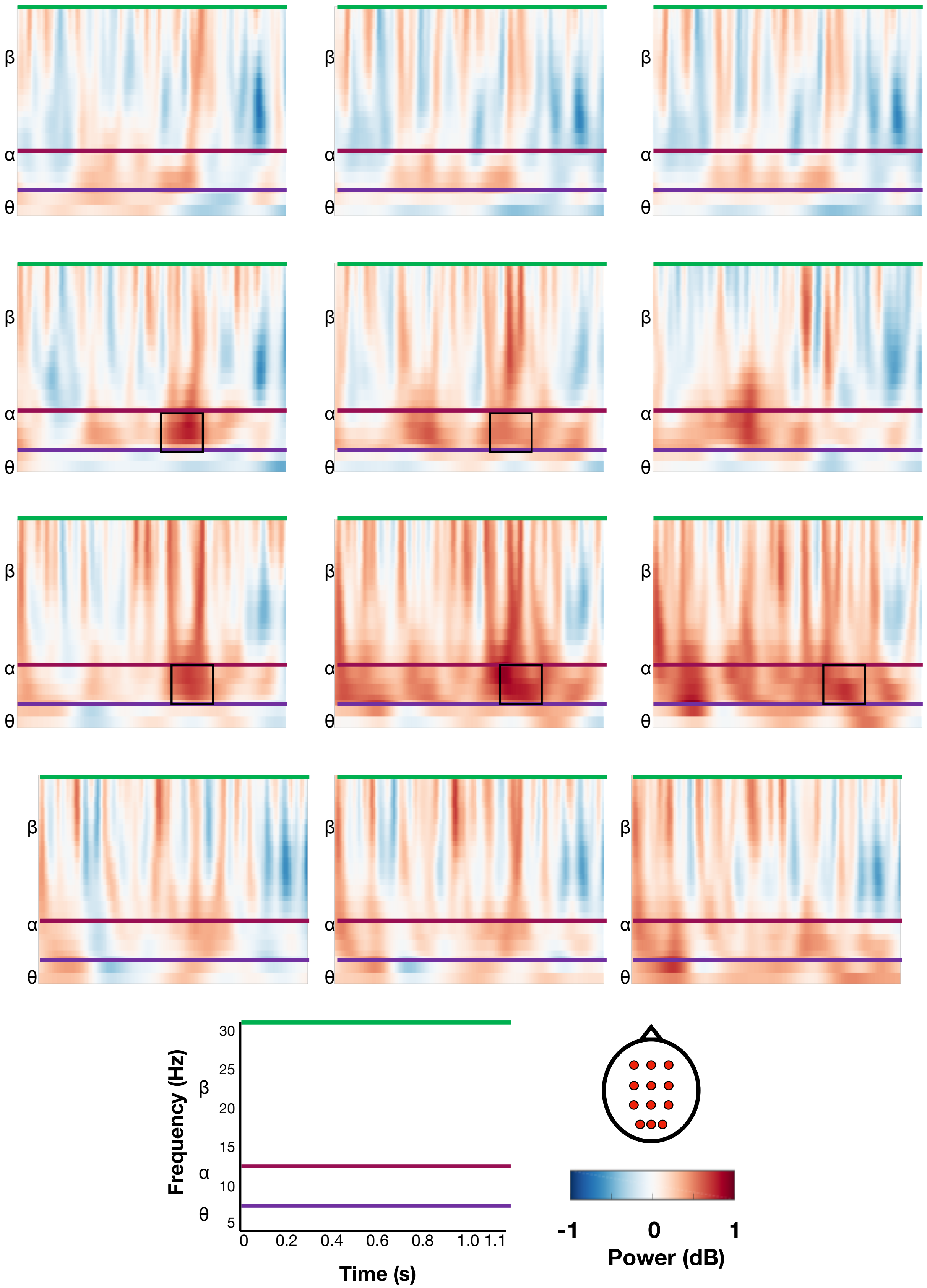

Figure 6.17. Time-Frequency representations for Test 1 bias. Plots denote power differences between future close and far bias conditions in Test 1, plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections denote time, electrode and frequency data part of a significant cluster.

For Test 2, the main effect of repetition bias showed significant clusters in both theta and alpha frequencies. These positive clusters corresponded to increased theta/alpha power for close repetition versus far repetition bias conditions from approximately 0.2-0.7s at left parietal and right frontal electrodes (see Figure 6.18). No significant clusters were found for the interaction of accuracy and repetition bias. In summary, these results show increased theta/alpha power was observed during selection of targets and distractors when participants repeated a similar response to previous retrieval attempts.

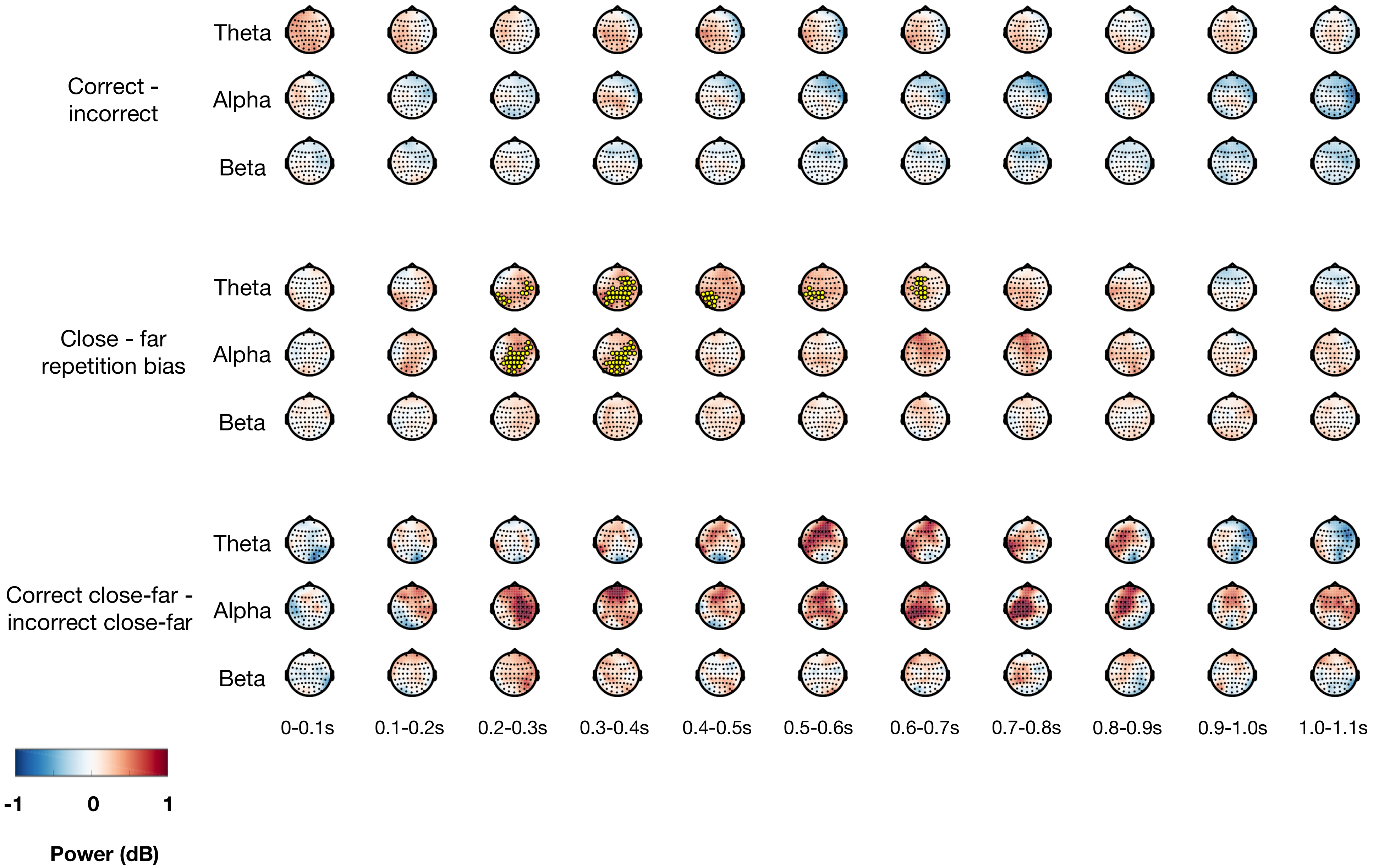

Figure 6.18. Cluster plots for oscillatory analysis of test 2 accuracy and bias. Topographical maps with significant clusters (denoted by yellow dots) demonstrating main effects of accuracy and bias, as well as an accuracy-bias interaction in test 2 for theta, alpha and beta bandwidths. Only the largest clusters are plotted.

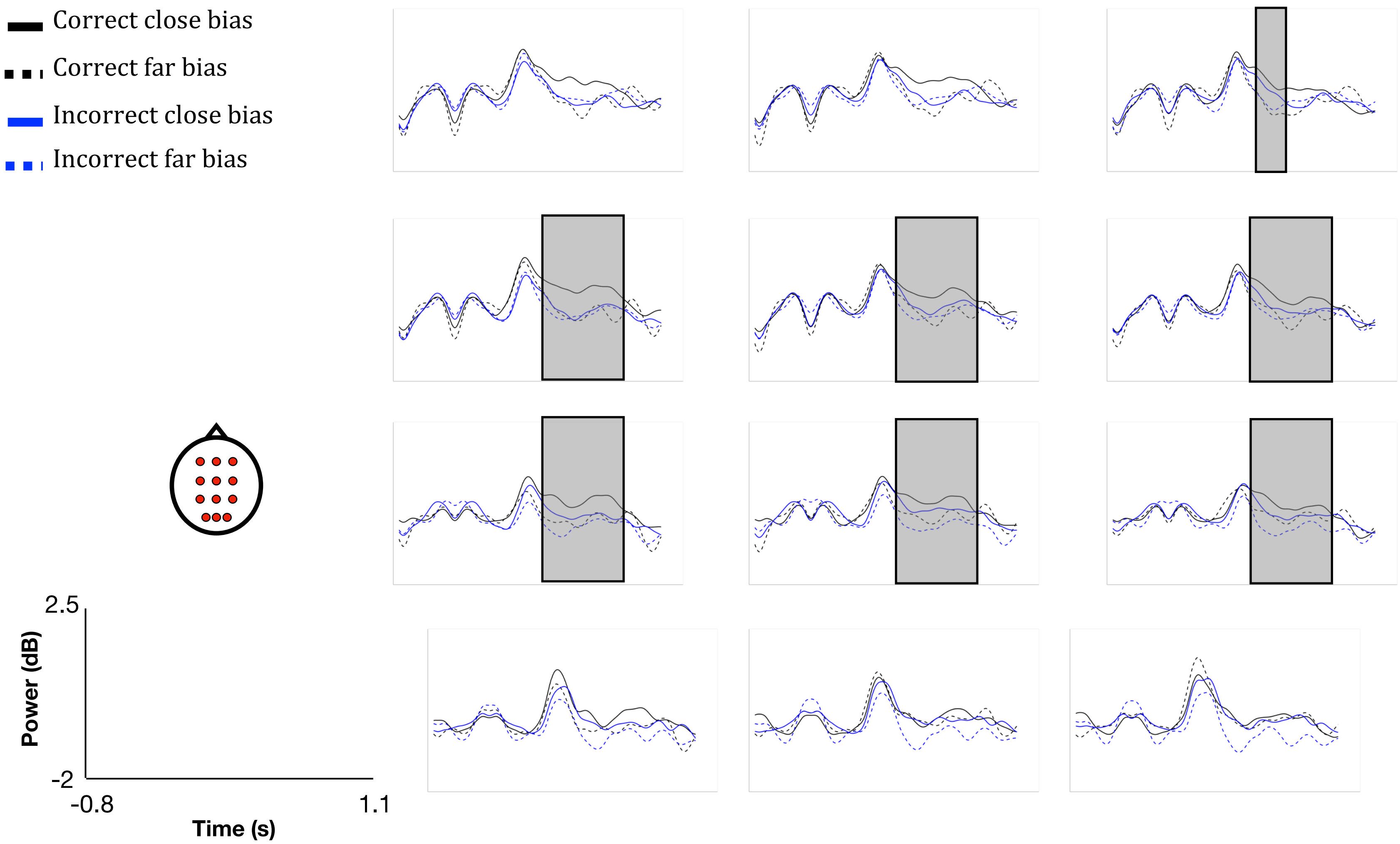

Figure 6.19. Time-frequency plots of theta power for test 2 accuracy and bias. Line plots of theta power for correct close, correct far, incorrect close and incorrect far conditions in test 2. Theta power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections represent significant time points.

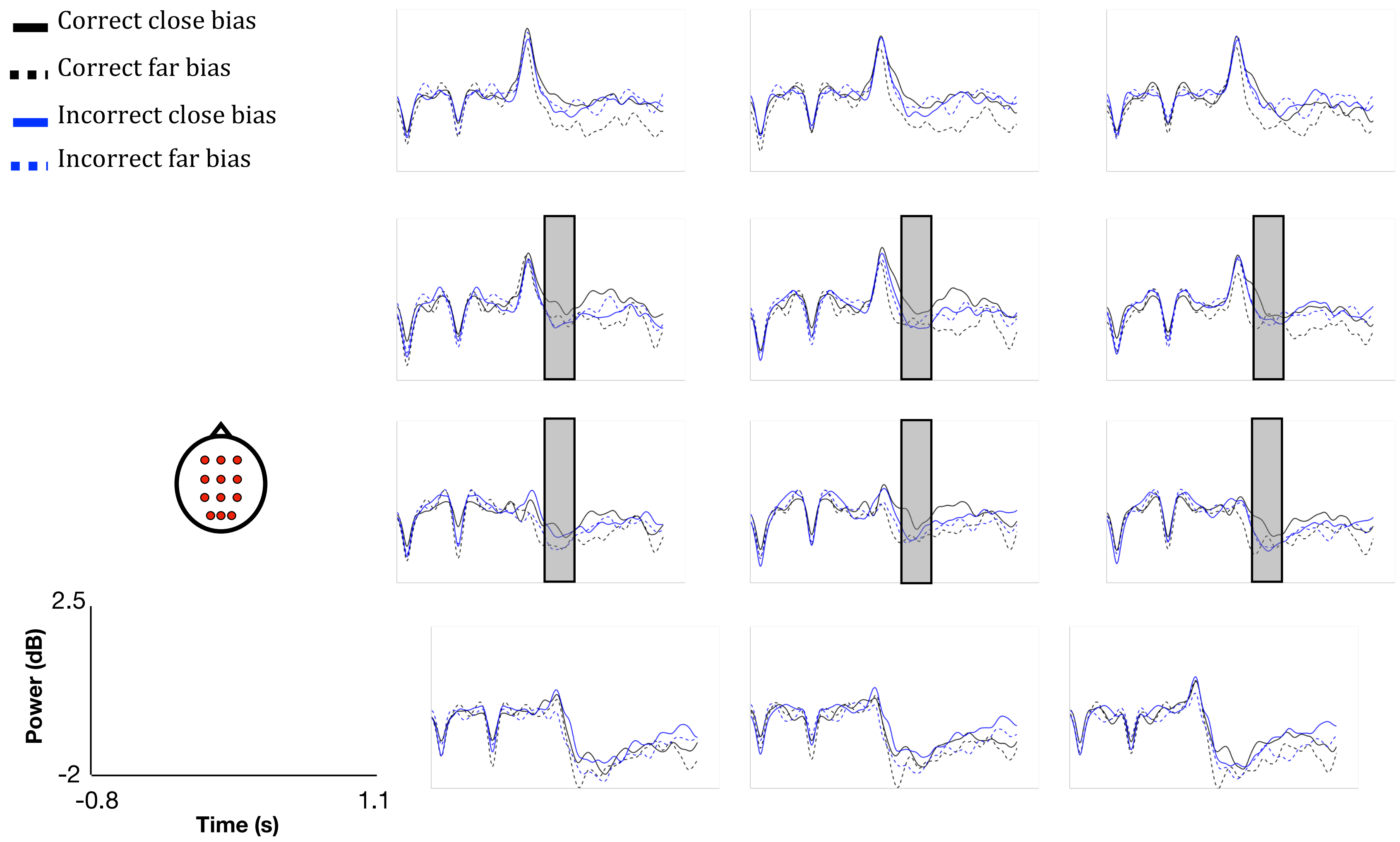

Figure 6.20. Time-frequency plots of alpha power for test 2 accuracy and bias. Line plots of theta power for correct close, correct far, incorrect close and incorrect far conditions in test 2. Alpha power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections represent significant time points.

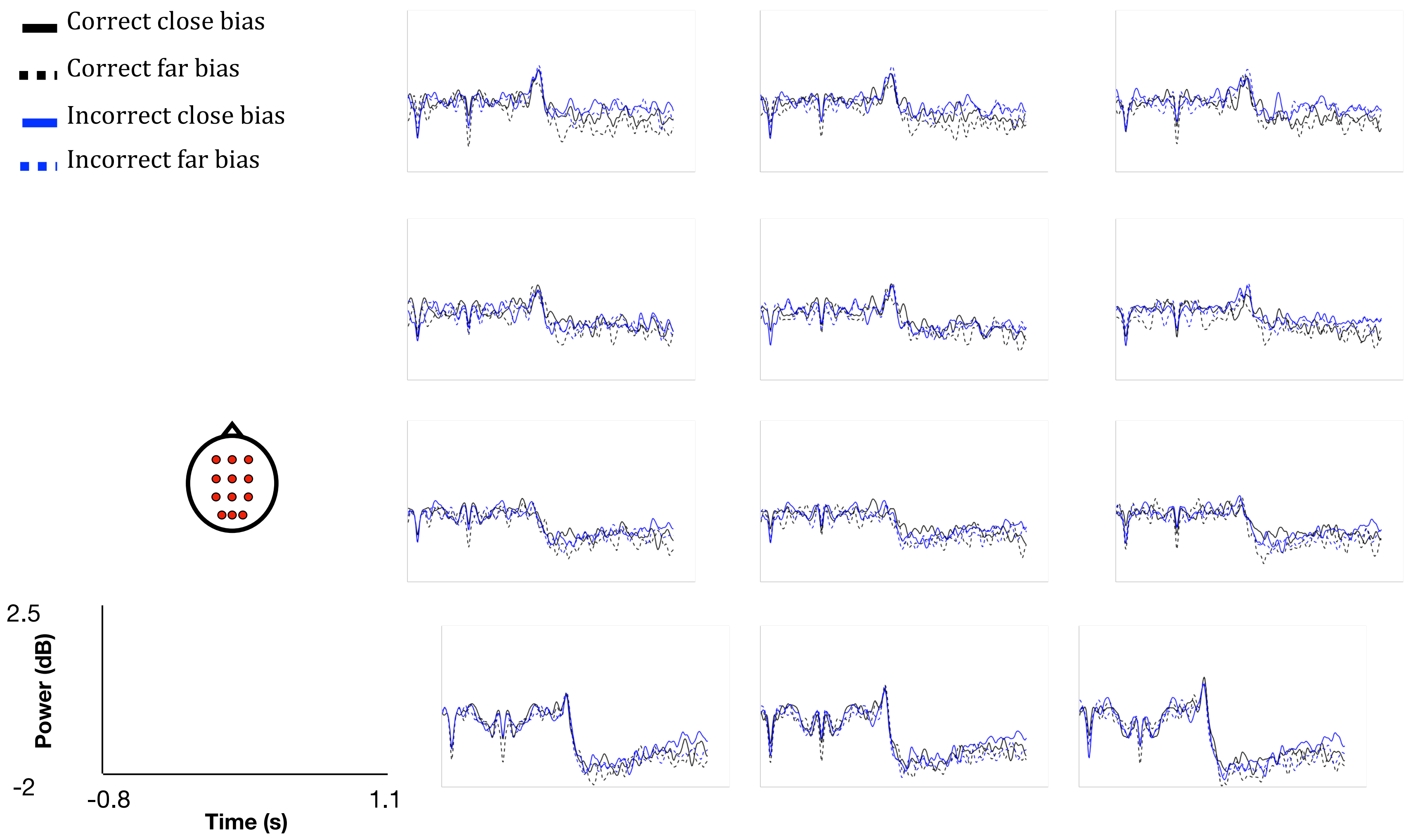

Figure 6.21. Time-frequency plots of beta power for test 2 accuracy and bias. Line plots of theta power for correct close, correct far, incorrect close and incorrect far conditions in test 2. Beta power plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2.

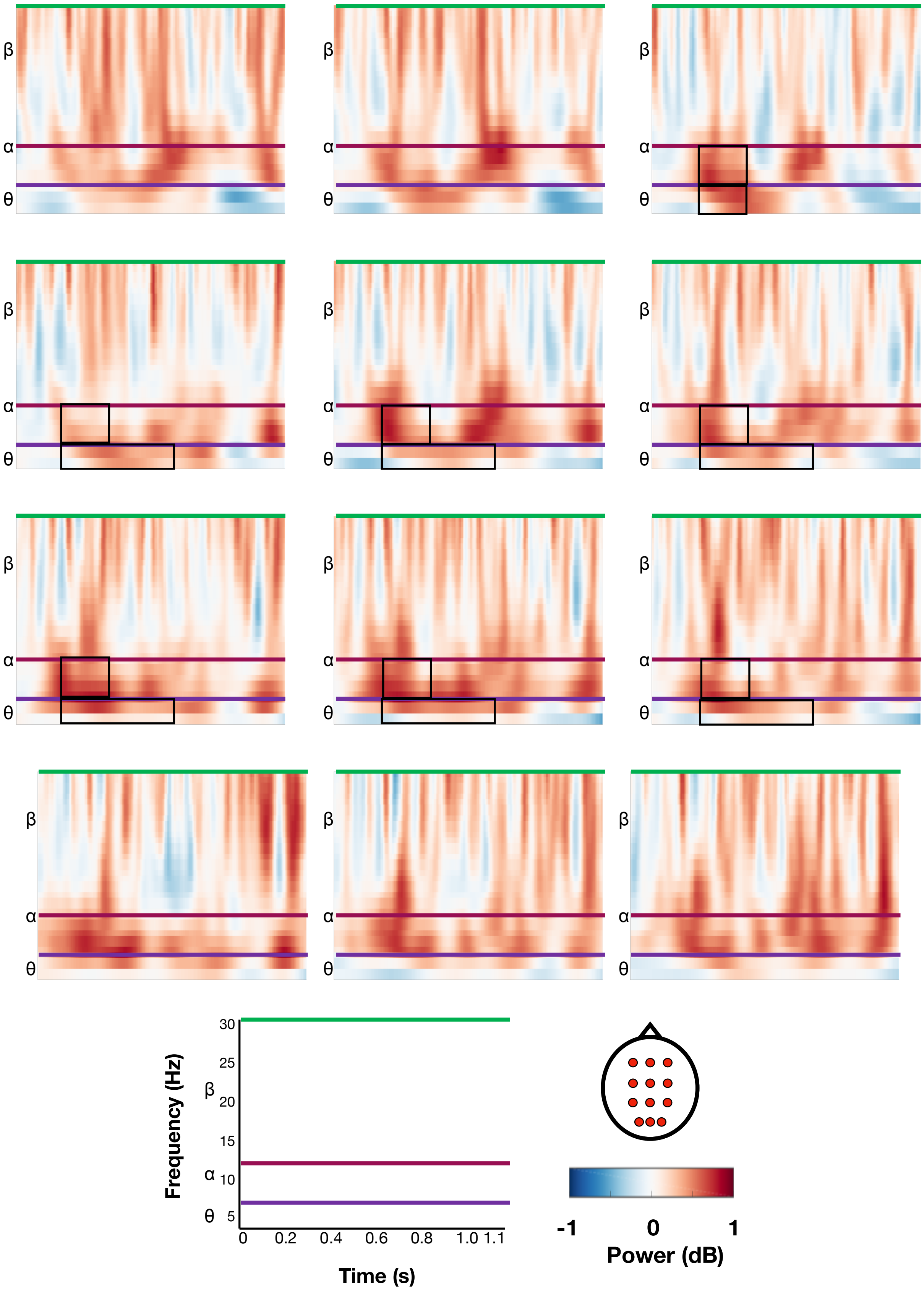

Figure 6.22. Time-Frequency representations for Test 2 bias. Plots denote power differences between repetition close and far bias conditions in Test 2, plotted at F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz and O2. Shaded sections denote time, electrode and frequency data part of a significant cluster.

6.3. Discussion

The aim of the analyses in Chapter 6 was to examine the oscillatory correlates of objective recognition accuracy, subjective retrieval experience and memory updating mechanisms during repeated face recognition. EEG data from Experiment 5 were used to estimate oscillations, for which behavioural and ERP results were presented in Chapter 5. During Test 1, late beta desynchronisations were shown to correlate with correct compared to incorrect recognition. However, for both accurate and erroneous retrieval attempts, high compared to low confidence judgements were associated with alpha/beta desynchronisation. In Test 2 the same effect was shown for high confidence recognition responses for both correct and incorrect recognition, yet alpha/beta oscillations no longer distinguished correct from incorrect responses. Also during Test 1, alpha synchronisation was associated with correct recognition responses that would be repeated during the subsequent Test 2 recognition task. During Test 2 recognition, theta/alpha synchronisation was found to correlate with retrieval of face memories that were similar to previous recognition responses, both when these responses were accurate and erroneous. These effects in Test 1 and 2 can be identified as the oscillatory mechanisms associated with the retrieval-induced updating of face memories. Finally, between test analysis established that, theta synchronisation effects for correct compared to incorrect recognition attempts was significantly different and numerically reversed in direction, between Test 2 than test 1.

The alpha/beta desynchronisation effects for high confidence judgements during Test 1 and Test 2 can be interpreted as an oscillatory correlate of subjective retrieval processes. In both tests, the alpha/beta desynchronisation effect was observed for high confidence recognition responses when both targets and distractors were selected. As a reminder, faces in the incorrect high conditions were, on average, perceptually dissimilar to target face images, thus alpha/beta power decreases for these conditions were not due to these conditions containing faces that were very similar to target images. These results thus show that alpha/beta desynchronisation correlated with subjective experience of face recognition, regardless of whether those faces had objectively been seen before.

Previous work has suggested that alpha/beta oscillations reflect an inhibitory process in the brain (Klimesch et al., 2007; Jensen & Mazaheri, 2010). Thus desynchronisation of these oscillations suggests a ‘release of inhibition’ in material-specific regions of the brain relevant for the task (Khader et al., 2010; Waldhauser, Johansson & Hanslmayr, 2012). The alpha/beta desynchronisation effects shown during Test 1 and 2 were primarily located at centro-posterior locations, with effects emerging towards the end of the epoch. Potentially, the posterior location of the alpha/beta effect in the present analysis may reflect a disinhibition of visual regions of the brain that contribute to episodic memory retrieval, such as the precuneus (Richter et al., 2016), or regions critical for the processing of faces, including the fusiform face gyrus (Loffler et al., 2005; Kanwisher & Yovel, 2006; Kriegeskorte, 2007) and occipital face area (Pitcher et al., 2009; Pitcher et al., 2011). As mentioned in Chapter 5, electrophysiological correlates of face processing are thought to occur early (e.g. N170 & N250), yet the claims made from the current experiment relate to much later mechanisms of subjective recognition experience as opposed to early detection and reactivation of face memories. The late onset of the beta desynchronisation effect occurs after the typical timings of electrophysiological correlates of conscious episodic memory retrieval (e.g. the FN400 and LPE; Rugg & Curran, 2007; Wilding & Ranganath, 2011). However, the functional properties of what this effect could relate to, in terms of subjective retrieval processing of face memories, is unclear.

In Test 1, alpha/beta synchronisation effects were shown to correlate with correct recognition responses during Test 1, only if participants went on to make a similar recognition response in Test 2. One functional mechanism proposed for alpha/beta oscillations suggests that these oscillations inhibit neural activity (Klimesch et al., 2007; Jensen & Mazaheri, 2010). Therefore, one possible but tentative suggestion could be that an increase in alpha/beta power during correct recognition of target face memories resulted in the engagement of mechanisms to inhibit processing of interfering information, such as distractor faces. That is, within a given trial, face stimuli were designed to vary in degrees of similarity to target faces. Thus, when participants recognised target face images in a given trial, it is possible that inhibitory mechanisms were engaged to prevent interference from competing distractor faces.

This argument is consistent with previous studies showing that, during memory encoding, alpha/beta power increases are observed during the maintenance of items in short-term/working memory (Bonnefond & Jensen, 2012; Khader et al., 2010; Meeuwissen et al., 2010; Poch et al., 2018; Samuel et al., 2018). Additionally, a study by Bäuml et al. (2008) showed increased alpha power during presentation of a word list predicted whether participants would successfully recall words from this list at the expense of words from a list learnt previously. Furthermore, in a study by Park et al. (2014), increased alpha activity was found for items that were “to-be-ignored” during memory encoding, and this alpha effect was source localised to the parietal cortex in the dorsal attentional network. Consequently, in the present study, it is possible that the inhibition of distractor processing during target recognition served to prioritise processing of target memories, which could facilitate their re-encoding to aid the successful retrieval of these memories again during subsequent remembering.

A possible mechanism of the brain inhibiting competing representations to prioritise re-encoding of target faces assumes that participants experienced a competition/conflict between the recognised target and distractor images within, or even between trials, possibly due to the similarity of images that led to multiple faces being recognised (as indicated in the ‘multiple memory’ data in Experiment 3a, see Section 3.1.2). This suggestion, however, needs to be elaborated on in future work as the EEG experiment did not ask participants if they experienced conflicting recognition of multiple faces during Test 1 or 2. By comparing trials where participants experienced retrieval conflict to trials without retrieval conflict, changes in alpha power could be investigated to assess if alpha increases reflect the need to inhibit competing information in order to prioritise the processing of to-be-remembered information.

As mentioned, alpha synchronisation during Test 1 predicted whether a similar recognition response would be given during test 2. Alpha oscillations are thought to reflect a general inhibitory mechanism that facilitates cognition but does not necessarily reflect a core memory process. Thus, an oscillatory correlate of memory updating that correspond to memory-specific functions was not observed. Consistent with previous literature (Blumenfeld & Ranganath, 2007), theta synchronisation effects may have been expected during the encoding of face representations that updated memories. Theta power synchronisation is thought to facilitate episodic memory formation by synchronising activity between medial temporal lobe and neocortical brain regions, a network that is critical for distant regions of the brain to communicate during memory formation (Nyhus & Curran, 2010; Parish et al., 2018). Indeed, during Test 1, a cluster corresponding to a theta power increase was observed for memories that would be repeated during Test 2, however this cluster did not meet the alpha threshold so cannot be interpreted as supporting this prediction. In future studies, research could consider enhancing the updating effects observed in the experiments of this thesis, as mentioned in the discussion section of Chapter 6. For example, updating could be enhanced by presenting face images for longer time periods (such as in Pezdek et al., 2005). If updating can be enhanced by behavioural manipulations during the repeated recognition task, then encoding-related neural correlates may be more likely to emerge.

One final recommendation for future research would be to consider whether the similarity of distractor faces recognised during Test 1 influences the neural mechanism of updating. That is, behavioural data from Experiment 5 (replicating results from Experiments 3a & b) showed that during Test 2, participants were more likely to select similar faces to those selected in Test 1 if these faces were more similar to the target. These results emphasise how recognition errors in this paradigm can vary significantly in terms of similarity to the target, therefore it is possible that different neural mechanisms may correlate with the updating of face memories by false information that is similar to original memories, vs. information that is dissimilar to the original memory. For example it has already been mentioned in the discussion of Chapter 3 (in Section 3.3) that cognitive mechanisms of pattern separation (Yassa & Stark, 2010) may determine whether false information that is similar to the target modifies these memories (i.e. failures to pattern separate; e.g. Li et al., 2019; Sun et al., 2017) or creates a separate memory trace to spare the original memory from modification (i.e. successful pattern separation). Such analyses would require an increased number of trials than those acquired from the current experiment, to ensure that enough trials could be allocated to conditions where recognised faces during initial retrieval were similar or dissimilar to target faces, and whether these faces were selected again during subsequent recognition.

A further consideration as to why no significant encoding-related updating effects emerged during Test 1 can be attributed to the nature of the statistical test used. Cluster-based permutation tests were used to analyse the statistical significance of ERP/time-frequency effects. The use of this technique, compared to more standard analytical tools such as ANOVA's, was warranted for the present thesis for several reasons (see Section 2.4.4 for more detail). However, one natural consequence of cluster-based permutation test is that the cluster significance is dependent on the size of the cluster itself, with longer lasting differences that span a wider network of electrodes more likely to be detected using the cluster correction method. Thus, smaller differences between conditions may exist but not be detected using the cluster corrected permutation test. Nevertheless, the absence of significant effects should not be attributed to just the type of statistical test used. Instead it is important to verify whether a meaningful effect is present first of all by adopting the described changes from the previous paragraph in future studies.

A final point to acknowledge with the oscillatory data in this chapter considers the signal-to-noise ratio of this analysis. This issue, addressed in previous chapters, is important when considering the 10-trial cut-off that was adopted for participant inclusion in the analysis of Experiment 5. It is possible that, should more trials have been allocated to conditions for each participant, there may have been an increased signal-to-noise ratio to detect meaningful modulations of oscillatory power between conditions. This is an important issue for the interaction analyses (interaction effects both within and between tests) where noise is doubled for each interaction 'subtraction' that is performed. Consequently, future adaptations of the present work should consider a higher minimum trial number for each participant, enhancing the signal-to-noise ratio and statistical power of oscillatory analyses.

The next finding showed that, during Test 2, left parietal to right-frontal theta and alpha synchronisation was seen during recognition of faces that were similar to recognition responses made in Test 1. Critically, such theta/alpha synchrony was not dependent on whether the selected face was correct (the target) or incorrect (a distractor) during Test 2. Given that, on average, incorrect recognition responses were perceptually dissimilar to target face representations, the finding that theta/alpha synchrony was shown for incorrect recognition responses (despite not reaching the alpha threshold) may suggest that participants were retrieving a modified episodic memory that was updated by encoding processes during the preceding Test 1 task. Supporting this argument, previous research has established theta power synchronisation correlates with successful episodic memory retrieval (Addante et al., 2011; Hanslmayr et al., 2011, Wynn et al., 2019). The present findings thus suggest that, during a repeated recognition task, theta/alpha synchronisation across left-parietal and right-frontal regions supports the retrieval of original and updated episodic memories that have been strengthened/modified by previous recognition attempts.

The final consideration of the reported analyses considers the between test comparisons of the accuracy-confidence conditions. It was shown that theta synchronisation effects were larger for correct compared to incorrect retrieval attempts in Test 2 compared to Test 1. As mentioned, theta power has been associated with successful episodic memory retrieval, with frontal theta effects linked to control mechanisms during retrieval (Klimesch et al., 1997; Waldhauser et al., 2012), whilst theta power increases at parietal electrodes have been associated with recollection judgements (Addante et al., 2011; Hanslmayr et al., 2011, Wynn et al., 2019). The fact that theta power increases were observed for correct recognition attempts in Test 2 rather than Test 1 may suggest that, during Test 2, participants engaged in a more controlled retrieval search and recollected specific aspects of the learning context associated with face memories, to recognise targets. Such controlled retrieval processes may not have been necessary for Test 1 as target faces could only have originated from the learning phase at this time. Therefore, participants may have been able to rely on more automatic forms of retrieval processing to distinguish targets from distractors during the first test than the second test.

In summary, the final experimental chapter of the current thesis examined the oscillatory correlates of the objective, subjective and updating-related processes during face recognition. The critical findings demonstrated that; 1) posterior alpha/beta desynchronisation correlated with enhanced subjective retrieval experience, regardless of objective retrieval accuracy, 2) alpha/beta synchrony during initial retrieval attempts predicted future recognition responses when retrieval was consistently accurate, and 3) theta/alpha synchronisation during repeated retrieval occurred when participants were repeating similar recognition responses to those in a previous test, including both accurate and erroneous recognition decisions, thus showing evidence of the retrieval of memories updated by previous recognition attempts.

Chapter 7 – Summary, conclusions and future research

7.1. Summary of the thesis

The aim of this thesis was to examine the cognitive and neurocognitive mechanisms underlying the retrieval-induced updating of face memories. It is known that episodic memories can become updated when a neural trace representing a memory is 'reactivated' (Gershmann et al., 2013; Hardt et al., 2010; Hupbach et al., 2007; 2008; 2009; Sederberg et al., 2011). Furthermore, prior work has established that episodic memories can become updated by information that is encoded during retrieval attempts (Bridge & Paller, 2012; Liu et al., 2018), with updating more likely when memories are remembered with an elevated subjective retrieval experience (St. Jacques et al., 2013; St. Jacques & Schacter, 2013). Finally, episodic memory updating has been shown to be enhanced following retrieval of information from memory vs. re-studying the same information (Bridge & Voss, 2014). However, prior to this thesis, much less was known regarding the neurocognitive mechanisms underlying the updating of simpler episodic memories. This topic was investigated in this thesis by studying the updating of face memories from repeated recognition attempts.

To summarise the results across the five experimental chapters, each experiment presented face images derived from artificial face space models (Solomon et al., 2013) in a novel repeated recognition paradigm (see Appendix A for face image verification experiments). In this task, participants attempted to recognise target faces (shown during a previous learning phase) shown amongst four distractor images, in two separate tests. In Chapter 3, increased confidence in judgements during the first recognition test predicted that participants would select the same face during subsequent recognition. This effect that was observed when both targets (i.e. correct recognition) and distractors (i.e. incorrect recognition) were selected during test 1. In the next chapter, Experiments 4a-c demonstrated that face memory updating was increased following active recognition attempts compared to re-study tasks that required participants to encode faces that were cued (Experiment 4a & c) or selected based on a distinctiveness judgement (Experiment 4c). Finally, the electrophysiological correlates of face memory retrieval and updating were examined with ERPs (Chapter 5) and oscillations (Chapter 6). Results in Chapter 5 showed that late positive ERPs distinguished the recognition of targets vs. distractors, as well as faces recognised with high vs. low confidence, an effect most likely reflecting reactivation of face memories. In Chapter 6, oscillatory EEG effects also covaried with objective and subjective retrieval processes (alpha/beta desynchronisation), however oscillatory correlates of updating were also shown. In test 1, alpha oscillations were associated with recognition of targets that would subsequently also be recognised during test 2. However, during test 2, theta/alpha oscillations were enhanced during recognition of faces that were similar to prior selections, for both correct and incorrect responses, thus this effect may be a possible correlate of retrieval of updated face memories.

The findings across all experiments in this thesis are the first to provide evidence of the neurocognitive mechanisms underlying the recognition-induced updating of face memories. To reiterate, the two key findings from behavioural data were; 1) that confidence of initial recognition attempts predicts subsequent memory updating, and 2) retrieval vs. re-study enhances face memory updating. The confidence effects were demonstrated across three separate experiments (Experiment 3a, 3b & 5), despite all three experiments having a slightly varied task design (e.g., categorical vs. continuous confidence scales, number of items per block, different number of learning cycles), showing the robustness of this effect. Furthermore, the behavioural data from these three experiments demonstrated a consistent finding that, during initial retrieval, selected faces that closely resembled target memories were more likely to be selected again during subsequent recognition. This finding was possible to show due to the benefit of using continuous recognition measures, that being the Euclidean distance between faces in an artificial image space.

Within Experiments 3a-b and Experiment 5, the suggested neurocognitive processes underlying face memory updating were largely based on previous literature on memory updating (Bridge & Paller, 2012; Liu et al., 2018; St.Jacques et al., 2013). These studies argue that two processes occur during retrieval that may lead to the updating of retrieved memories. Firstly, the neural trace of stored memories is reactivated during retrieval, after which encoding mechanisms are engaged to modify the reactivated memories with information present during the retrieval environment. The same framework could also account for the updating of face memories being correlated with high confidence judgements, such that face memories could be reactivated during recognition (possibly enhanced for memories remembered with increased confidence, e.g. St. Jacques et al., 2013), with encoding mechanisms storing selected faces into long-term memory (also possibly enhanced by increased confidence by increased attention to selected faces, e.g. Ciaramelli et al., 2008).

The EEG data from Experiment 5 were analysed to verify these suggestions, with ERP and oscillatory correlates of the repeated face recognition paradigm analysed. The ERP data mainly showed effects relating to the reactivation of face memories recognised with high confidence or repeatedly correctly recognised (c.f. Mackenzie et al., 2018; Mackenzie & Donaldson, 2007; 2009), with no evidence of ERPs relating to the encoding of faces selected during initial recognition that predicted subsequent updating. In contrast, oscillatory markers of memory updating were found, with alpha oscillations during test 1 predicting repeated correct recognition, and theta oscillations during test 2 correlating with repeated recognition of targets and distractors. However, these analyses did not find a neural correlate during false recognition attempts at initial retrieval that led to subsequent retrieval distortions (i.e., no ERP or oscillatory markers were observed for incorrect recognition attempts in Test 1 that led to a close bias response in Test 2), an aspect that can be explored in future research recommendations (see Section 7.2).

The electrophysiological results from this thesis provide a unique insight into the neurocognitive mechanisms underlying face memory retrieval and updating, due to the differences found between Chapters 5 and 6. As a reminder, ERPs generally corresponded to reactivation associated with retrieval of targets, and recognition of targets and distractors associated with high confidence. In contrast, oscillatory mechanisms reflected markers of both reactivation (in particular for high confidence recognition judgements) and updating. The differences of ERP and oscillatory analyses may reflect evoked vs. induced neural processes associated with face memory retrieval and updating. That is, ERPs are thought to reflect the averaged, evoked activity in response to a stimulus that is consistent across trials. Time-frequency analysis, however, decomposes EEG into underlying frequency components reflecting induced activity that will vary in their onset/amplitude in response to a stimulus across trials (David et al., 2006; Bastiaansen et al., 2011). The functional distinctions between evoked and induced neural processes, such that evoked processes may reflect bottom-up processes whilst induced processes may reflect top-down mechanisms (Chen et al., 2012; Tallon-Baudry & Bertrand, 1999), has important considerations for interpreting the present EEG findings.

If ERPs are thought to reflect bottom-up processes evoked from sensory processing of stimuli, then the ERP findings from Chapter 5 may reflect an evoked mechanism of reactivating memory traces associated with the recognition of faces, as all faces in the current study may have evoked some reactivation during recognition (given the similarity of face stimuli, and the presentation of the same face images throughout the experiment). The ERP effects in Chapter 5 had temporal onsets of ~0.4-0.7s and lasted until the end of the epoch (1.1s), time points consistent with the onset of familiarity and recollection-related recognition processes (Mackenzie & Donaldson, 2007, 2009; Mackenzie et al., 2018; Rugg & Curran, 2007), as well as with suggested time courses of hippocampal-induced pattern completion processes associated with retrieval (Staresina & Wimber, 2019; though no strong claims can be made regarding the involvement of specific brain regions with the current EEG data set). However, the temporal onset of specific retrieval-related processes (such as familiarity and recollection) could vary across trials (that is, participants may not always recollect details of a memory at the same time across trials). Therefore, it could be argued that the ERP effects observed in this thesis speak towards an evoked mechanism of reactivating neural traces associated with face memories.

Furthermore, the lack of ERP correlates of updating processes may be due to the varied temporal onset that these processes would be engaged/induced during retrieval. For example, inhibiting competing information during target recognition (in the form of an alpha power increase) would not necessarily be evoked in response to target face presentation, but would occur as soon as participants require the need to inhibit competing information and maintain attentional resources towards target faces. Additionally, the late alpha/beta desynchronisation effects could reflect an induced top-down process where perceptual processes are reinstated (as in a feedback mechanism) following the initial reactivation and retrieval of face memories (Staresina & Wimber, 2019) to facilitate recognition judgements. Finally, the theta/alpha power increases during repeated recognition, suggested to correlate with retrieval of updated face memory representations, would be induced as and when these memories are retrieved. Consequently, the present electrophysiological data from this thesis demonstrate the utility of assessing both the ERP and oscillatory mechanisms of face memory recognition and updating. These findings emphasise the importance of assessing both ERPs and oscillations in episodic memory research as a crucial approach to fully understand the mechanistic processes underlying memory retrieval and updating.

The mechanisms of memory reactivation and encoding were also used to account for the data from experiments in Chapter 4. In these experiments, it was shown that retrieval of face memories, compared to re-study tasks that required the explicit encoding of faces, enhanced the repeated selection of distractor faces at subsequent recognition. The conclusions from this Chapter were strengthened by the multi-experimental approach taken to address the issue of retrieval vs. re-study on face memory updating, with several potential confounding factors (such as a block order effect, or a ‘self-choice’ effect) ruled out as explanations for the increase in biased responses during the final recognition task following retrieval vs. re-study tasks. As discussed in Chapter 4 (see Section 4.4), the neurocognitive mechanisms underlying enhanced memory updating following retrieval vs. re-study could be attributed to either a ‘reactivation’ mechanism (increased reactivation of memories during retrieval vs. re-study) or an ‘encoding’ mechanism (faces selected during retrieval receive more encoding than faces selected during re-study). These ideas need to be verified in future work using the paradigm from Experiments 4a-c with neuroimaging techniques (see Section 7.2).

Across all of the experiments in this thesis, face memory updating was operationalised as when false recognition/selection of distracting information led to these distractors being recognised again during subsequent retrieval attempts, at the expense of recognising target faces that were learnt prior to the presentation of distractors. Each of the behavioural experiments also examined the extent to which accurate recognition of targets changed across repeated retrieval attempts. The consistent pattern of findings across experiments in Chapters 3 and 4 (Experiment 4a & b) demonstrated that general accuracy rates were highest during initial retrieval, with a similar decline in accuracy during a subsequent retrieval test for both memories that underwent retrieval during prior testing (i.e. repeated conditions) and memories that did not undergo prior testing (i.e. baseline conditions). One exception to this was from Experiment 5 (see Section 5.2), where proportion accuracy for the repeated condition was significantly larger than the baseline condition (though only an accuracy difference of 4% was observed). The pattern of results whereby recognition accuracy was similar for the repeated and baseline conditions is inconsistent with predictions from retrieval practice paradigms which would suggest higher recognition accuracy for repeated vs. baseline conditions during Test 2 (Roediger & Butler, 2011). As mentioned in these chapters, the way in which recognition accuracy is conditionalised (on all Test 1 trials, or just correct Test 1 trials), and the level of performance during initial retrieval (Rowland & Delosh, 2014; Smith et al., 2013) can alter whether retrieval practice benefits are observed. The results from these chapters contribute to this literature by showing how retrieval accuracy for face memories is influenced by the repeated recognition paradigms used in this thesis.

A relevant point regarding the choice and justification of methods/analyses in this thesis considers the use of receiver operater characteristic (ROC) curves. ROC curves plot the relationship between hit rates (i.e. correctly recognising targets) and false alarm rates (i.e. incorrectly recognising a distractor face) as a function of different response criterions (Urquhart & O’Connor, 2018; Yonelinas & Parks, 2007), such as varying levels of recognition confidence (Van Zandt, 2000). ROC curves are then fitted to model parameters that represent single or dual-process models of recognition memory (Juola et al., 2009; Yonelinas & Parks, 2007). Thus, whilst ROC curves could have been used to model recognition performance in the current experiments, this thesis was less concerned with the underlying processes associated with recognition memory (e.g. familiarity vs. recollection) but more concerned with how varying levels of confidence modulate face memory updating. However, familiarity vs recollection related processes may have important implications for updating of face memories, given that memory reactivation may be more likely during recollection compared to familiarity, possibly increasing the likelihood of memory updating (Leiker & Johnson, 2014; Johnson et al., 2009), thus these issues could be addressed in future experiments to understand the underlying recognition processes associated with face memory updating.

The tasks used in the experiments of this thesis were designed with the assumption that participants were following and paying attention to the instructions set out from the experimenter. This is particularly valid for the experiments in Chapter 4, where each experiment modulated the refresh task, thus participant compliance with the instructions of these tasks was important for the validity of the refresh manipulations. However, as already acknowledged, participants may have ignored the instructions for the re-study and selection tasks and could have prioritised attentional resources to the face they recognised from the learning phase. Without asking participants on a trial by trial basis as to whether they recognised any faces within a given a trial, it is impossible to know to what extent this occurred in the present data (a potential change that could be implemented in future experiments). However, such concerns were addressed with the bias difference score which measured the extent to which participants were more likely to select the same distractor, compared to switching to a different distractor, during final recognition. As seen in the re-study and select refresh conditions of Experiments 4a and 4c, participants were more likely to repeatedly select the same distractor more than selecting a different distractor during final recognition, implying that participants were encoding distractor faces to some degree in these two conditions.

For the experiments in Chapters 3 and 5, participants were instructed to select faces they recognised from the original learning phase, and to provide confidence judgements in their recognition responses. It is difficult to argue of an alternate process/task that participants could have been adopting in these experiments, although participants could have focused on selecting the same face they picked from the preceding recognition task, rather than constraining their retrieval memory search to the learning phase. This suggestion is consistent with the view that participants could have simply forgotten target faces by test 1, instead encoding a distractor face during test 1 which was repeated in test 2. However, if this was the case, then it would be expected to see larger bias scores for low rather than high confidence responses (as participants would have low confidence if they did not recognise any of the faces). Instead, bias measures were reduced following low vs. high confidence responses when distractor faces were falsely recognised, suggesting that participants were, at least generally, complying with task instructions to remember faces according to their memory from initial learning tasks.

A final methodological consideration of this thesis questions the statistical power of each experiment. Statistical power has become an important consideration of psychological and cognitive neuroscience studies in recent years, owing to claims of a replication crisis in such fields that may, in part, be due to underpowered studies (Button et al., 2013; Szucs & Ioannidis, 2017). To attempt to ensure that research is sufficiently powered, researchers can determine a sample size ahead of data collection that is able to detect an effect size of interest with a high level of statistical power. However, in this thesis, each experiment was conducted without a priori power calculations to determine a sample size in part because the true effect size for the contrasts of interest would be difficult to estimate (due to the novelty of the stimuli and paradigm presented). Instead, sample sizes in the present thesis were largely determined to recruit as many participants as possible within time-limited recruitment windows, with sample sizes roughly guided by those often used in episodic memory research.

In respecting the issue of statistical power, for each experiment, power analyses were conducted to assess the power of each experiment to detect both medium (0.5) and small (0.3) effect sizes. Generally, each experiment showed power values lower than the widely used benchmark of 80% (i.e., 80% chance to detect an effect when a true difference exists in the population), particularly to detect small effect sizes, which is important to consider for the current thesis as the majority of effect sizes for the key updating contrasts were small. Therefore, some of results in this thesis, particularly for non-significant results and for contrasts with small effect sizes, must be viewed with a degree of caution due to a reduced statistical power to detect a difference if these differences existed in the population. Nonetheless, the results should not be disregarded completely, instead future studies should build on the work in this thesis to demonstrate the replicability of these findings, with sufficiently powered experiments.

From a theoretical perspective, the results from this thesis can be interpreted with respect to two prominent models to account for the updating of episodic memories; reconsolidation (Dudai & Eisenberg, 2004; Hardt et al., 2010) and the Temporal Context Model (Howard & Kahana, 2002; Sederberg et al., 2011). Reconsolidation suggests that episodic memories are modified after memory traces are reactivated and converted from a stable, dormant trace to an active, labile representation. During this time-dependent memory reactivation, memories can become modified by information present in the retrieval environment, with the modified trace ‘reconsolidated’ into a dormant form that may be remembered at subsequent retrieval. Applied to the present thesis, it is possible that face memories that were reactivated (during test 1 in Experiments 3a-b, or retrieval/re-study tasks in Experiments 4a-c) were modified based on what face was selected during these tasks. Reactivated memories would then be modified according to whether selected faces were targets (leading to target memory representations being strengthened) or distractors (leading to target memory representations being distorted).

Some behavioural evidence from Experiments 3a-b and 5 lends support to reconsolidation theory, specifically, with the finding that participants were more likely to repeat a similar face selection during test 2 when the face selected in test 1 was more similar to target faces. In this scenario, the target representation may have been distorted by perceptually similar distractors, creating a ‘blurred’ representation that distorted subsequent recognition attempts (e.g. Li et al., 2019; Sun et al., 2017). An important point here may constrain the reconsolidation model, such that target memories would only be reconsolidated by distractor faces if these images were perceptually similar to the target, with the brain failing to detect fine-grained perceptual differences between these types of information to create separate traces between the original target memory and false distractor face memories (e.g., via pattern separation; Yassa & Stark, 2010).

However, face memory updating can also occur when distractor faces are perceptually distinct from target faces, (evidenced in Chapter 5 with the mean Euclidean distance between target faces and the incorrect distractor selections). Whilst it is possible that more perceptually distinct distractor faces may also distort the representation of target face memories via reconsolidation, it is less clear why these faces would be selected having not been perceived previously. Instead, it seems more plausible that perceptually distinct distractors would be selected following sub-optimal encoding of target memories. For these reasons, it is entirely possible that several neurocognitive mechanisms of memory updating occur and depend on several factors including the quality of target face encoding, and the similarity of targets to distractors faces selected during initial retrieval/re-study attempts.

Whilst reconsolidation is a prominent theory of memory updating, several constraints of this theory may question its applicability to the data from this thesis. Reconsolidation theory makes clear predictions that the reactivation and reconsolidation of modified memories occurs across time scales of several hours (via synaptic consolidation; Bramham & Messaoudi, 2005) to even days and weeks (via systems consolidation; Dudai, 2004), time scales that are much longer than the timings used in the present thesis where encoding and retrieval tasks were separated by 15-20 minutes. Possibly, the reconsolidation theory needs to be reconsidered to account for the changes in memory representations that occur within minutes of encoding, retrieval and re-encoding. Furthermore, an issue with the theory of reconsolidation is that it is difficult to conclusively prove that a memory representation in the human brain is directly modified. Supporting this point, data from Experiment 3a showed that participants reported an increase in recognising multiple faces as a reason for providing low confidence judgements during the second recognition task (see Section 3.1.2). Therefore, updating may occur not by modifying stored representations in the brain, but instead by a retrieval competition of separate memory traces; the original target memory and a memory formed during retrieval/re-study tasks (e.g. McCloskey & Zaragoza, 1985; Riccio, Millin & Bogart, 2006).

An alternative theory of memory updating, the Temporal Context Model (Howard & Kahana, 2002) proposed by Sederberg et al. (2011) suggests that reactivation occurs for the temporal context of an episodic memory. The reactivated temporal context then becomes bound with information present during reactivation (e.g. faces selected during retrieval/re-study tasks). Such a mechanism may have occurred in the experiments within this thesis, as the temporal context associated with the learning phase may be reactivated during initial retrieval/re-study tasks, meaning that faces selected during these tasks could have been bound to this reactivated context. Should distractors be bound to the learning temporal context, participants may have then have misattributed distractor faces as being from the original learning phase. This model may be able to better account for the findings in Experiment 4, as it would be expected that remembering (thus reactivating) the temporal context of the learning phase would only be necessary during retrieval where participants had to remember which face was shown during learning, compared to re-study tasks that required participants to encode faces during this task without having to remember or reactivate the learning context.

An important point to raise when considering the theoretical explanations of face memory updating needs to consider that reconsolidation and Temporal Context Model theories are largely derived from tasks where participants were encoding and retrieving lists of items, and whether participants would misattribute items to alternative lists. In contrast, the paradigms presented in this thesis focused on updating of face memories at an item-level. Therefore, to advance the understanding of how item memories become updated, future work could look to adapt memory updating models to account for the updating of item memories. One way to advance this research is to consider models of face recognition, such as the face space often referred to throughout this thesis (Lewis, 2004; Valentine, 2001; Valentine et al., 2015). Such work could adapt current models of face space (e.g. Lewis, 2004), to consider how faces are represented in the brain (Loffler et al., 2005), whilst also introducing elements that account for face memory updating effects postulated in this thesis. This approach could address several issues, such as explaining how, for a given face that has a varied level of encoding error (as already modelled by Lewis, 2004), how would the neural representation of this face be modulated according to whether participants recognised a different face within (or beyond) the error boundary.

The discussion above raises the issue of how the results from this thesis compare to previous episodic memory updating findings. A major rationale for the research conducted in this thesis was that previous research had focused on investigating the neurocognitive mechanisms of updating of elaborate episodic memories, such as the updating of items associated with contextual information via cued-recall (Bridge & Paller, 2012; Bridge & Voss, 2014; Liu et al., 2018) or the updating of rich, autobiographical memories (St. Jacques et al., 2013; St. Jacques & Schacter, 2013). Therefore, this thesis aimed to examine the cognitive and neurocognitive mechanisms of much simpler forms of episodic memories, that being the recognition of faces. For several reasons, face recognition may involve a different set of mechanisms compared to updating of more elaborate episodic memories. For example, because memory updating may be enhanced by reconstructive processes (St. Jacques et al., 2013), updating of item memories via familiarity-based recognition may be less likely or reduced compared to retrieval-induced updating of elaborate memories via recollection (Leiker & Johnson, 2014; Johnson et al., 2009). Nevertheless, the results from this thesis did have some resemblance to previous findings, such that increased subjective retrieval processes were associated with memory updating (e.g. St. Jacques et al., 2013) and active retrieval attempts vs. re-study boosted memory updating (e.g. Bridge & Voss, 2014). Furthermore, the ERP data from Experiment 5 were similar to those seen previously (Bridge & Paller, 2012; Liu et al., 2018). It is possible that both item recognition and item-context recall engage the same mechanisms, but the content of what is being reactivated and encoded would differ according to the content of retrieved memories and the type of retrieval task. For example, for the recognition tasks in the experiments of this thesis, the reactivation of the temporal context associated with face memories may occur, whereas reactivation of context in the cued-recall tasks of Bridge & Paller (2012) and Bridge & Voss (2014), for example, may instead involve spatial contextual information. Therefore, more work is necessary to understand the mechanistic similarities and differences between updating of simpler vs. elaborate episodic memories.

A further contribution of this thesis suggests some factors that may contribute to why eyewitnesses may repeatedly select the same face across multiple identification attempts. As reviewed in the introductory chapter, the ‘commitment effect’ has shown that eyewitnesses may be likely to select the same face across multiple identification attempts following a crime, even when the person of recognised faces is not the actual suspect (see Steblay & Dysart, 2016). The results from this thesis could provide useful information in the design of repeated identification procedures with eyewitnesses. In particular, the role of confidence during identification attempts has been studied and debated with respect to whether high confidence judgements during identification attempts can be used as a proxy to determine successful suspect identification. Indeed, it has been argued that an eyewitness could be trusted to have correctly identified a suspect more likely with high vs. low confidence judgements (Wixted et al., 2018, so long as the police follow recommended procedures, e.g. a fair line-up, avoid contaminating an eyewitness memory). Research studying the role of eyewitness confidence in relation to repeated identification procedures has measured eyewitnesses confidence during initial and repeated identifications. Studies have found that witnesses were just as confident for correct and incorrect identification attempts during initial (Steblay et al., 2013) and repeated identification attempts (Goodsell et al., 2009; Steblay et al., 2013). However, less research has been done to examine whether confidence judgements during initial identification attempts predict whether witnesses would select the same face in a subsequent identification attempt. In this thesis, it was also found that participants were recognising distractor images with high confidence that predicted repeated incorrect selections at future recognition attempts. Consequently, the data from the experiments of this thesis could contribute to a wider discussion of the role of eyewitness confidence predicting successful suspect identification in repeated identification procedures.

In order for the results of this thesis to apply to eyewitness procedures, it is important to consider whether the artificial face images used in this thesis adequately capture the variability of real human face images, and whether such face memories are applicable to real eyewitness contexts (where face memories are associated with multiple features such as suspect body/clothing, emotional valence of the criminal event, etc). Regarding the face stimuli, the external features of the artificial faces (ears, neck & hair) were blurred, meaning that only the internal facial features varied between images. Additionally, the same pictures of individual faces were represented throughout the experiment, thus it could be argued that these facial images fail to capture the natural variability of faces (Burton, Kramer, Ritchie & Jenkins, 2016). However, the use of these images was justified to achieve the general goal for the current thesis to examine the mechanisms of face memory updating whilst controlling for several extraneous variables and to be able to use continuous metrics of recognition performance, an aspect that contributed to various results presented throughout the thesis. Nevertheless, this approach limits the applicability of the present results to repeated identification attempts in police investigations. Future research could build on the results of the present thesis by using photographs of real human faces, with more realistic eyewitness procedures (time delay between encoding and retrieval of face memories, emotional valence of encoding events, associating faces with full bodies) to examine whether the factors considered in the thesis predict repeated eyewitness identification attempts.

As mentioned, future work can build on the findings from this thesis to examine factors that predict repeated eyewitness identification attempts. One proposed change to the experimental paradigm, having more realistic time delays between encoding and retrieval of targets faces, raises a relevant question as to whether updating would be observed at a longer time delay between the encoding and retrieval of faces that have never been seen prior to the tasks. To elaborate, with the paradigm used in this thesis, memory for target faces may not have transformed from unstable to consolidated memory traces as there were only a few minutes between encoding and recognition tasks (Bramham & Messaoudi, 2005; Dudai, 2004). Therefore, it is possible that the effects observed in this thesis correspond to the learning and development of face representations whilst the memories are in an initial state of instability, given that a face representation develops by repeatedly perceiving several pictorial variants of the same facial identity (Burton et al., 2015). Nevertheless, the question of how a facial identity develops vs. how a memory for picture of a face can be updated are two distinct set of questions, therefore an idea for future research may assess the updating of face memories that have undergone a sustained period of initial consolidation.

7.2. Future research

Whilst this thesis has contributed to our understanding of neurocognitive mechanisms underlying face memory updating, there are several questions that remain unanswered, none more so than clarifying the neural mechanisms underlying face memory updating. As mentioned in Chapters 5 and 6, no neural correlate was found during the initial false recognition of distractors that predicted subsequent memory updating, despite very clear predictions that encoding mechanisms during initial retrieval tasks must be engaged for distractors to distort target face memories. One could argue, given the absence of any neural correlates of encoding-during-retrieval in test 1, whether encoding mechanisms were engaged during initial retrieval to promote memory updating. This would challenge the argument made throughout this thesis arguing that factors, such as confidence (Chapter 3, 5 & 6) or retrieval vs. re-study (Chapter 4), could modulate encoding of faces to update memories. However, it is more plausible to argue that encoding processes are engaged that promote the updating of face memories (for example, following high vs low confidence recognition attempts), yet encoding processes are also present for distractor faces recognised with low confidence (given that they were completely novel during test 1), minimising the differences in encoding processes between high and low confidence responses.

One change to the experimental paradigms of this thesis to build on the above issue is to enhance the memory bias effect by enhancing the encoding of faces selected during initial retrieval/refresh tasks. This could be achieved by presenting face images for longer durations (e.g. Pezdek et al., 2005), by reducing the number of faces per set, or by making face images more distinct from each other, to enhance encoding of selected faces during these tasks. It would be expected that enhancing the encoding of face images would boost the extent to which selected faces can update memories, increasing the likelihood of observing neural correlates of updating during initial retrieval/re-study tasks. Alternatively, to address the potential argument that encoding mechanisms were boosted for novel distractor faces (even if they were recognised with low confidence), the repeated recognition task could be amended to have three or more (rather than two) recognition tests. Analysis of updating mechanisms could then be addressed in the second test (after all faces have been presented but before a subsequent third recognition task).

Nevertheless, it may the case that the predicted mechanisms of updating based on previous research may not necessarily apply to updating effects observed in this thesis. Specifically, episodic memory updating mechanisms are thought to be dependent on the hippocampus (Bridge & Voss, 2014; St. Jacques et al., 2013), whilst the oscillatory mechanisms of theta and gamma encoding and retrieval are also linked to hippocampus (Parish et al., 2018). There has been a discussion questioning whether the recognition of unfamiliar faces (i.e., recognition of faces not seen prior to the experiment) is governed by the hippocampus (Bird, 2017), especially with shorter delays between the encoding and recognition of faces. This issue is highly relevant to the paradigms of this thesis, as faces had never been perceived prior to the experiment (therefore are ‘unfamiliar’) with the delay between learning and recognition tasks being a relatively short delay of 5-10 minutes. Therefore, in order to further understand the neural mechanisms of face memory updating, including the verification of the proposed mechanisms of reactivation and encoding-during-retrieval, and how they compare to the updating mechanisms of other types of episodic memory, further work needs to be done to understand which brain regions are associated with face memory retrieval and updating, using techniques with enhanced spatial resolution such as fMRI.

Also mentioned previously in this discussion section, it is important to advance on the current thesis by studying the nature of how face memory representations are stored, modified and updated in the brain through repeated retrieval attempts. A lot of the discussion in this thesis has taken assumptions from face space models to explain how representations of faces *may* become updated throughout repeated retrieval. In particular, the nature of how memory representations of faces were stored and updated was based on assumptions regarding the ‘encoding error’ of face targets. Thus, understanding the neural representations of a face memories in the brain is integral to the understanding of how these memories can become modified. In recent years, neuroimaging techniques have advanced for us to answer questions regarding the representational nature of episodic memories. Pattern classification techniques such as MVPA (Norman et al., 2006) and RSA (Kriegeskorte et al., 2008) analyse patterns of brain activity to assess whether distinct patterns of activity represent certain types of information in the brain, such as the content of retrieval (e.g. Lee et al., 2018; Staresina et al., 2012), and even patterns of activity that distinguish individual faces during perception (Kriegeskorte et al., 2007; Lee & Kuhl, 2016) and memory (Lee & Kuhl, 2016).

Pattern classification techniques therefore have the potential to answer some of the questions regarding the representational nature of face memory updating. Nevertheless, several caveats must be noted and accounted for in future research. Firstly, pattern classification techniques in episodic memory research are most often applied to cued or free recall tasks where remembered information that is retrieved with or without cues is classified, content from which can only be derived from an internally generated retrieval search. In contrast, pattern classification techniques are not often applied to recognition tasks for the reason that classification of brain activity during recognition would not be able to distinguish between activity from the perception of information during recognition, from the activity related to the retrieval of episodic memories (e.g. Chadwick et al., 2010; Weil & Rees, 2010). For this reason, the paradigms in the current thesis would have to be adapted when using pattern classification techniques to assess how memory representations become updated.

A second issue to raise regarding the use of pattern classification techniques with the current paradigm is whether the faces used in this thesis are distinct enough to decode unique activity patterns for each of the individual faces. Previous studies that have used pattern classification where the face images differed on distinct perceptual dimensions such as gender (Kriegeskorte et al., 2007; Lee & Kuhl, 2016) and race (Lee & Kuhl, 2016). In contrast, the face images in this thesis were derived from Caucasian faces, with dimensions such as gender or age not specified or varied, a constraint of the principal components analysis procedure that creates the artificial face space (see Section 2.1). Therefore, the question remains whether unique neural patterns for these very similar face images can be decodable. Whilst some promising work has shown that overlapping episodic memories can be successfully decoded (e.g. Bonnici et al., 2011; Chadwick, Hassabis & Maguire, 2011), evidence of successful decoding of perceptually similar faces is yet to be established and can be an avenue of future research to consider.

7.3. Conclusion

To conclude, this thesis is the first to provide evidence of the cognitive and neurocognitive mechanisms underlying the updating of face memories, with retrieval confidence and retrieval practice two factors shown to modulate face memory updating. The proposed mechanisms by which faces become modified by repeated retrieval attempts were addressed using ERP and oscillatory data, yet future work is needed to further specify the neural mechanisms by which face representations become modified through repeated retrieval. Future work on this topic should bring together computational models of face recognition with neurocognitive models of episodic memory to answer the questions regarding the mechanistic nature of face memory updating. This work has practical implications for eyewitness testimony procedures by beginning to provide an understanding as to why eyewitness memory for a suspect's face may become updated/distorted through repeated identification attempts.

References

Addante, R. J., Ranganath, C., & Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: evidence of accurate source recognition without recollection. *NeuroImage, 62(1),* 439–450. doi:10.1016/j.neuroimage.2012.04.031.

Addante, R. J., Watrous, A. J., Yonelinas, A. P., Ekstrom, A. D., & Ranganath, C. (2011). Prestimulus theta activity predicts correct source memory retrieval. *Proceedings of the National Academy of Sciences, 108(26),* 10702–10707. doi:10.1073/pnas.1014528108.

Ally, B. A., Simons, J. S., McKeever, J. D., Peers, P. V., & Budson, A. E. (2008). Parietal contributions to recollection: electrophysiological evidence from aging and patients with parietal lesions. *Neuropsychologia, 46(7),* 1800–1812. doi:10.1016/j.neuropsychologia.2008.02.026.

Andrews, S., Burton, A. M., Schweinberger, S. R., & Wiese, H. (2017). Event-related potentials reveal the development of stable face representations from natural variability. *Quarterly Journal of Experimental Psychology, 70(8),* 1620–1632. doi:10.1080/17470218.2016.1195851.

Axmacher, N., Henseler, M. M., Jensen, O., Weinreich, I., Elger, C. E., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences, 107(7),* 3228–3233. doi:10.1073/pnas.0911531107.

Babu Henry Samuel, I., Wang, C., Hu, Z., & Ding, M. (2018). The frequency of alpha oscillations: Task-dependent modulation and its functional significance. *NeuroImage, 183*, 897–906. doi:10.1016/j.neuroimage.2018.08.063.

Bastiaansen, M., Mazaheri, A., & Jensen, O. (2011). Beyond ERPs: oscillatory neuronal dynamics. In E. S. Kappenman, & S. J. Luck (Eds.). *The Oxford Handbook of Event-Related Potential Components.* New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780195374148.013.0024.

Bäuml, K.-H., Hanslmayr, S., Pastötter, B., & Klimesch, W. (2008). Oscillatory correlates of intentional updating in episodic memory. *NeuroImage, 41(2),* 596–604. doi:10.1016/j.neuroimage.2008.02.053.

Bell, A. J. & Sejnowski, T. J. (1995). An information maximisation approach to blind separation and blind deconvolution. *Neural Computation, 7(6),* 1129-1159. doi:10.1162/neco.1995.7.6.1129.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8(6),* 551–565. doi:10.1162/jocn.1996.8.6.551.

Bergström, Z. M., Henson, R. N., Taylor, J. R., & Simons, J. S. (2013). Multimodal imaging reveals the spatiotemporal dynamics of recollection. *NeuroImage, 68,* 141–153. doi:10.1016/j.neuroimage.2012.11.030.

Bird, C. M. (2017). The role of the hippocampus in recognition memory. *Cortex*, 93, 155–165. doi:10.1016/j.cortex.2017.05.016.

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(4),* 657–668. doi:10.1037/0278-7393.15.4.657

Blumenfeld, R. S., & Ranganath, C. (2006). Dorsolateral prefrontal cortex promotes long-term memory formation through its role in working memory organization. *Journal of Neuroscience, 26(3),* 916–925. doi:10.1523/jneurosci.2353-05.2006.

Blumenfeld, R. S., & Ranganath, C. (2007). Prefrontal cortex and long-term memory encoding: an integrative review of findings from neuropsychology and neuroimaging. *The Neuroscientist, 13(3),* 280–291. doi:10.1177/1073858407299290.

Bonnefond, M., & Jensen, O. (2012). Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Current Biology, 22(20)*, 1969–1974. doi:10.1016/j.cub.2012.08.029.

Bonnici, H. M., Kumaran, D., Chadwick, M. J., Weiskopf, N., Hassabis, D., & Maguire, E. A. (2011). Decoding representations of scenes in the medial temporal lobes. *Hippocampus, 22(5),* 1143–1153. doi:10.1002/hipo.20960.

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2017). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology, 55(6),* e13049. doi:10.1111/psyp.13049.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11(5),* 164–169. doi:10.1111/1467-8721.00192.

Bramham, C. R., & Messaoudi, E. (2005). BDNF function in adult synaptic plasticity: The synaptic consolidation hypothesis. *Progress in Neurobiology, 76(2),* 99–125. doi:10.1016/j.pneurobio.2005.06.003.

Bridge, D. J., & Paller, K. A. (2012). Neural correlates of reactivation and retrieval-induced distortion. *Journal of Neuroscience, 32(35),* 12144–12151. doi:10.1523/jneurosci.1378-12.2012.

Bridge, D. J., & Voss, J. L. (2014). Hippocampal binding of novel information with dominant memory traces can support both memory stability and change. *Journal of Neuroscience, 34(6),* 2203–2213. doi:10.1523/jneurosci.3819-13.2014.

Bridger, E. K., Bader, R., Kriukova, O., Unger, K., & Mecklinger, A. (2012). The FN400 is functionally distinct from the N400. *NeuroImage, 63(3),* 1334–1342. doi:10.1016/j.neuroimage.2012.07.047.

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2015). Identity from variation: representations of faces derived from multiple instances. *Cognitive Science, 40(1),* 202–223. doi:10.1111/cogs.12231.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7(1),* 26–48. doi:10.3758/bf03210724.

Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L. (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology, 20(7),* 941–956. doi:10.1002/acp.1239.

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., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature Reviews Neuroscience, 9(8),* 613–625. doi:10.1038/nrn2459.

Capelo, A. M., Albuquerque, P. B., & Cadavid, S. (2018). Exploring the role of context on the existing evidence for reconsolidation of episodic memory. *Memory, 27(3),* 280–294. doi:10.1080/09658211.2018.1507040.

Chadwick, M. J., Hassabis, D., & Maguire, E. A. (2011). Decoding overlapping memories in the medial temporal lobes using high-resolution fMRI. *Learning & Memory, 18(12),* 742–746. doi:10.1101/lm.023671.111.

Chadwick, M. J., Hassabis, D., Weiskopf, N., & Maguire, E. A. (2010). Decoding individual episodic memory traces in the human hippocampus. *Current Biology, 20(6)*, 544–547. doi:10.1016/j.cub.2010.01.053.

Chen, C.-C., Kiebel, S. J., Kilner, J. M., Ward, N. S., Stephan, K. E., Wang, W.-J., & Friston, K. J. (2012). A dynamic causal model for evoked and induced responses. *NeuroImage*, *59(1)*, 340–348. doi:10.1016/j.neuroimage.2011.07.066.

Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2008). Top-down and bottom-up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, *46(7)*, 1828–1851. doi:10.1016/j.neuropsychologia.2008.03.022.

Cohen, M. X. (2014). Analyzing neural time series data: theory and practice. *MIT Press, Cambridge.*

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.

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition, 28(6)*, 923–938. doi:10.3758/bf03209340.

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin, 136(1)*, 87–102. doi:10.1037/a0017937.

David, O., Kilner, J. M., & Friston, K. J. (2006). Mechanisms of evoked and induced responses in MEG/EEG. *NeuroImage, 31(4)*, 1580–1591. doi:10.1016/j.neuroimage.2006.02.034.

Davis, J. P., Gibson, S., & Solomon, C. (2014). The positive influence of creating a holistic facial composite on video line-up identification. *Applied Cognitive Psychology, 28(5)*, 634–639. doi:10.1002/acp.3045.

Davis, J. P., Maigut, A. C., Jolliffe, D., Gibson, S. J., & Solomon, C. J. (2015). Holistic facial composite creation and subsequent video line-up eyewitness identification paradigm. *Journal of Visualized Experiments, (106)*. doi:10.3791/53298.

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior, 30(3),* 287-307. doi:10.1007/s10979-006-9008-1.

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

DeSoto, K. A., & Roediger, H. L. (2014). Positive and negative correlations between confidence and accuracy for the same events in recognition of categorized lists. *Psychological Science, 25(3),* 781–788. doi:10.1177/0956797613516149.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2010). Medial temporal lobe activity during source retrieval reflects information type, not memory strength. *Journal of Cognitive Neuroscience, 22(8),* 1808–1818. doi:10.1162/jocn.2009.21335.

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

Dienes, Z. (2016). How bayes factors change scientific practice. *Journal of Mathematical Psychology, 72,* 78-89. doi:10.1016/j.jmp.2015.10.003.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology, 55(1),* 51–86. doi:10.1146/annurev.psych.55.090902.142050.

Dudai, Y., & Eisenberg, M. (2004). Rites of passage of the engram. *Neuron, 44(1),* 93–

100. doi:10.1016/j.neuron.2004.09.003.

Dudukovic, N. M., DuBrow, S., & Wagner, A. D. (2009). Attention during memory retrieval enhances future remembering. *Memory & Cognition, 37(7)*, 953–961. doi:10.3758/mc.37.7.953.

Eimer, M. (2011). *The Face-Sensitive N170 Component of the Event-Related Brain Potential*. Oxford Handbooks Online. doi:10.1093/oxfordhb/9780199559053.013.0017.

Friese, U., Köster, M., Hassler, U., Martens, U., Trujillo-Barreto, N., & Gruber, T. (2013). Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG. *NeuroImage, 66*, 642–647. doi:10.1016/j.neuroimage.2012.11.002.

Fuentemilla, L. (2018). Memory: theta rhythm couples periodic reactivation during memory retrieval. *Current Biology, 28(21)*, R1243–R1245. doi:10.1016/j.cub.2018.09.016.

Gershman, S. J., Schapiro, A. C., Hupbach, A., & Norman, K. A. (2013). Neural context reinstatement predicts memory misattribution. *Journal of Neuroscience, 33(20)*, 8590–8595. doi:10.1523/jneurosci.0096-13.2013

Gisquet-Verrier, P., Lynch, J. F., Cutolo, P., Toledano, D., Ulmen, A., Jasnow, A. M., & Riccio, D. C. (2015). Integration of new information with active memory accounts for retrograde amnesia: a challenge to the consolidation/reconsolidation hypothesis? T*he Journal of Neuroscience, 35(33)*, 11623–11633. doi:10.1523/jneurosci.1386-15.2015.

Gisquet-Verrier, P., & Riccio, D. C. (2012). Memory reactivation effects independent of reconsolidation. *Learning & Memory, 19(9)*, 401–409. doi:10.1101/lm.026054.112.

Goodsell, C. A., Neuschatz, J. S., & Gronlund, S. D. (2009). Effects of mugshot commitment on lineup performance in young and older adults. *Applied Cognitive Psychology, 23*, 788-803. doi:10.1002/acp.1512.

Graetz, S., Daume, J., Friese, U., & Gruber, T. (2018). Alterations in oscillatory cortical activity indicate changes in mnemonic processing during continuous item recognition. *Experimental Brain Research, 237(2)*, 573–583. doi:10.1007/s00221-018-5439-4.

Griffin, M., DeWolf, M., Keinath, A., Liu, X., & Reder, L. (2013). Identical versus conceptual repetition FN400 and parietal old/new ERP components occur during encoding and predict subsequent memory. *Brain Research, 1512*, 68–77. doi:10.1016/j.brainres.2013.03.014.

Guran, C.-N. A., Herweg, N. A., & Bunzeck, N. (2019). Age-related decreases in the retrieval practice effect directly relate to changes in alpha-beta oscillations. *The Journal of Neuroscience, 39(22)*, 4344–4352. doi:10.1523/jneurosci.2791-18.2019.

Hanslmayr, S., Spitzer, B., & Bauml, K.-H. (2008). Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cerebral Cortex, 19(7)*, 1631–1640. doi:10.1093/cercor/bhn197.

Hanslmayr, S., Staresina, B. P., & Bowman, H. (2016). oscillations and episodic memory: addressing the synchronization/desynchronization conundrum. *Trends in Neurosciences, 39(1)*, 16–25. doi:10.1016/j.tins.2015.11.004.

Hanslmayr, S., & Staudigl, T. (2014). How brain oscillations form memories — A processing based perspective on oscillatory subsequent memory effects. *NeuroImage, 85*, 648–655. doi:10.1016/j.neuroimage.2013.05.121.

Hanslmayr, S., Staudigl, T., & Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in Human Neuroscience, 6*. doi:10.3389/fnhum.2012.00074.

Hanslmayr, S., Volberg, G., Wimber, M., Raabe, M., Greenlee, M. W., & Bauml, K.-H. T. (2011). The relationship between brain oscillations and bold signal during memory formation: a combined eeg-fmri study. *Journal of Neuroscience, 31(44),* 15674–15680. doi:10.1523/jneurosci.3140-11.2011.

Hardt, O., Einarsson, E. Ö., & Nader, K. (2010). A bridge over troubled water: reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology, 61(1),* 141–167. doi:10.1146/annurev.psych.093008.100455.

Harlow, I. M., & Donaldson, D. I. (2013). Source accuracy data reveal the thresholded nature of human episodic memory. *Psychonomic Bulletin & Review, 20,* 318-325. doi:10.3758/s13423-012-0340-9.

Harlow, I. M., & Yonelinas, A. P. (2016). Distinguishing between the success and precision of recollection. *Memory, 24(1),* 114-127. doi:10.1080/09658211.2014.988162.

Haw, R. M., Dickinson, J. J., & Meissner, C. A. (2007). The phenomenology of carryover effects between show-up and line-up identification. *Memory, 15(1),* 117–127. doi:10.1080/09658210601171672.

Herrmann, M. J., Ehlis, A.-C., Ellgring, H., & Fallgatter, A. J. (2004). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). *Journal of Neural Transmission, 112(8),* 1073–1081. doi:10.1007/s00702-004-0250-8.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46(3),* 269–299. doi:10.1006/jmps.2001.1388.

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory, 14(1-2),* 47–53. doi:10.1101/lm.365707.

Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: updating or source confusion? *Memory, 17(5),* 502–510. doi:10.1080/09658210902882399.

Hupbach, A., Hardt, O., Gomez, R., & Nadel, L. (2008). The dynamics of memory: context-dependent updating. *Learning & Memory, 15(8),* 574–579. doi:10.1101/lm.1022308.

Itier, R. J. (2004). N170 or N1? spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex, 14(2),* 132–142. doi:10.1093/cercor/bhg111.

Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: theta correlates of memory retrieval and decision making. *NeuroImage, 32(2),* 978–987. doi:10.1016/j.neuroimage.2006.02.018.

Jarosz, A. F., & Wiley, J. (2014). what are the odds? a practical guide to computing and reporting bayes factors. *The Journal of Problem Solving, 7(1)*. doi:10.7771/1932-6246.1167.

JASP Team (2018). JASP (Version 0.8.6)[Computer software].

Jasper, H.H. (1958) The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371-375.

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

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Frontiers in Human Neuroscience, 4*. doi:10.3389/fnhum.2010.00186.

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. *Neuron, 63(5),* 697–708. doi:10.1016/j.neuron.2009.08.011.

Johnson, J. D., & Rugg, M. D. (2007). Recollection and the reinstatement of encoding-related cortical activity. *Cerebral Cortex, 17(11)*, 2507–2515. doi:10.1093/cercor/bhl156.

Juola, J. F., Caballero-Sanz, A., Muñoz-García, A. R., Botella, J., & Suero, M. (2019). Familiarity, recollection, and receiver-operating characteristic (ROC) curves in recognition memory. *Memory & Cognition, 47(4)*, 855–876. doi:10.3758/s13421-019-00922-8.

Kamp, S.-M., Bader, R., & Mecklinger, A. (2017). ERP subsequent memory effects differ between inter-item and unitization encoding tasks. *Frontiers in Human Neuroscience, 11*. doi:10.3389/fnhum.2017.00030.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361(1476)*, 2109–2128. doi:10.1098/rstb.2006.1934.

Kappenman, E. S., & Luck, S. J. (Eds.). (2011). *The Oxford Handbook of Event-Related Potential Components*. New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780195374148.001.0001.

Karpicke, J. D., & Blunt, J. R. (2011). retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331(6018)*, 772–775. doi:10.1126/science.1199327.

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62(3)*, 227–239. doi:10.1016/j.jml.2009.11.010.

Kempen, K., & Tredoux, C. G. (2012). “seeing is believing”: the effect of viewing and constructing a composite on identification performance. *South African Journal of*

Psychology, 42(3), 434–444. doi:10.1177/008124631204200315.

Kerrén, C., Linde-Domingo, J., Hanslmayr, S., & Wimber, M. (2018). An optimal oscillatory phase for pattern reactivation during memory retrieval. *Current Biology, 28(21)*, 3383–3392.e6. doi:10.1016/j.cub.2018.08.065.

Khader, P. H., & Rösler, F. (2011). EEG power changes reflect distinct mechanisms during long-term memory retrieval. *Psychophysiology, 48(3)*, 362–369. doi:10.1111/j.1469-8986.2010.01063.x.

Kim, H., & Cabeza, R. (2007). Trusting our memories: dissociating the neural correlates of confidence in veridical versus illusory memories. *Journal of Neuroscience, 27(45)*, 12190–12197. doi:10.1523/jneurosci.3408-07.2007.

Klimesch, W., Doppelmayr, M., Schimke, H., & Ripper, B. (1997). Theta synchronization and alpha desynchronization in a memory task. *Psychophysiology, 34(2)*, 169–176. doi:10.1111/j.1469-8986.1997.tb02128.x.

Klimesch, W., Doppelmayr, M., Schwaiger, J., Winkler, T., & Gruber, W. (2000). Theta oscillations and the ERP old/new effect: independent phenomena? *Clinical Neurophysiology, 111(5)*, 781–793. doi:10.1016/s1388-2457(00)00254-6.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews, 53(1)*, 63–88. doi:10.1016/j.brainresrev.2006.06.003.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language, 65(2)*, 85–97. doi:10.1016/j.jml.2011.04.002.

Köster, M., Friese, U., Schöne, B., Trujillo-Barreto, N., & Gruber, T. (2014). Theta–gamma coupling during episodic retrieval in the human EEG. *Brain Research, 1577*, 57–68. doi:10.1016/j.brainres.2014.06.028.

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences, 104(51),* 20600–20605. doi:10.1073/pnas.0705654104.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. doi:10.3389/neuro.06.004.2008.

Kuhl, B. A., & Chun, M. M. (2014). Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *Journal of Neuroscience, 34(23),* 8051–8060. doi:10.1523/jneurosci.4328-13.2014.

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

Lawson, V. Z., & Dysart, J. E. (2012). The showup identification procedure: An exploration of systematic biases. *Legal and Criminological Psychology, 19(1),* 54–68. doi:10.1111/j.2044-8333.2012.02057.x.

Leal, S. L., & Yassa, M. A. (2018). Integrating new findings and examining clinical applications of pattern separation. *Nature Neuroscience, 21(2),* 163–173. doi:10.1038/s41593-017-0065-1.

Lee, H., & Kuhl, B. A. (2016). Reconstructing perceived and retrieved faces from activity patterns in lateral parietal cortex. *The Journal of Neuroscience, 36(22),* 6069–6082. doi:10.1523/jneurosci.4286-15.2016.

Lee, H., Samide, R., Richter, F. R., & Kuhl, B. A. (2018). Decomposing parietal memory reactivation to predict consequences of remembering. *Cerebral Cortex, 29(8),* 3305–3318. doi:10.1093/cercor/bhy200.

Leiker, E. K., & Johnson, J. D. (2014). Neural reinstatement and the amount of information recollected. *Brain Research, 1582*, 125–138. doi:10.1016/j.brainres.2014.07.026.

Lega, B., Burke, J., Jacobs, J., & Kahana, M. J. (2014). Slow-Theta-to-gamma phase–amplitude coupling in human hippocampus supports the formation of new episodic memories. *Cerebral Cortex*, *26(1)*, 268–278. doi:10.1093/cercor/bhu232.

Lewis, M. (2004). Face-space-R: towards a unified account of face recognition. *Visual Cognition, 11(1)*, 29–69. doi:10.1080/13506280344000194.

Li, A. Y., Liang, J., Lee, A. C. H., & Barense, M. D. (2019). Visual interference can help and hinder memory: Measuring memory fidelity using a novel circular shape space. bioRxiv 535922; doi: https://doi.org/10.1101/535922.

Lisman, J. (2010). Working Memory: The Importance of theta and gamma oscillations. *Current Biology, 20(11)*, R490–R492. doi:10.1016/j.cub.2010.04.011.

Liu, X. L., Tan, D. H., & Reder, L. M. (2018). The two processes underlying the testing effect– Evidence from Event-Related Potentials (ERPs). *Neuropsychologia, 112*, 77–85. doi:10.1016/j.neuropsychologia.2018.02.022.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience, 8(10)*, 1386–1391. doi:10.1038/nn1538.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12(4)*, 361–366. doi:10.1101/lm.94705.

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

Mackenzie, G., Alexandrou, G., Hancock, P. J. B., & Donaldson, D. I. (2018), An item's status in semantic memory determines how it is recognized: Dissociable patterns of brain activity observed for famous and unfamiliar faces. *Neuropsychologia, 119,* 292-301. doi: 10.1016/j.neuropsychologia.2018.08.004

MacKenzie, G., & Donaldson, D. I. (2007). Dissociating recollection from familiarity: Electrophysiological evidence that familiarity for faces is associated with a posterior old/new effect. *NeuroImage, 36(2),* 454–463. doi:10.1016/j.neuroimage.2006.12.005.

MacKenzie, G., & Donaldson, D. I. (2009). Examining the neural basis of episodic memory: ERP evidence that faces are recollected differently from names. *Neuropsychologia, 47(13),* 2756–2765. doi:10.1016/j.neuropsychologia.2009.05.025.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164(1),* 177–190. doi:10.1016/j.jneumeth.2007.03.024.

Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14(2),* 194–199. doi:10.3758/bf03194051.

Mathworks. (2017). RST Toolbox (r2017a).

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102(3),* 419–457. doi:10.1037/0033-295x.102.3.419.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General, 114(1),* 1–16. doi:10.1037/0096-3445.114.1.1.

Meeuwissen, E. B., Takashima, A., Fernández, G., & Jensen, O. (2010). Increase in posterior alpha activity during rehearsal predicts successful long-term memory formation of word sequences. *Human Brain Mapping, 32(12),* 2045–2053. doi:10.1002/hbm.21167.

Mist, J. J., Gibson, S. J., & Solomon, C. J. (2015). Comparing evolutionary operators, search spaces, and evolutionary algorithms in the construction of facial composites. *Informatica, 39,* 135-145.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23(1),* 103-123. doi:10.3758/s13423-015-0947-8.

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.

Murray, J. G., Howie, C. A., & Donaldson, D. I. (2015). The neural mechanism underlying recollection is sensitive to the quality of episodic memory: Event related potentials reveal a some-or-none threshold. *NeuroImage, 120,* 298-308. doi:10.1016/j.neuroimage.2015.06.069.

Murty, V. P., DuBrow, S., & Davachi, L. (2015). The simple act of choosing influences declarative memory. *Journal of Neuroscience, 35(16),* 6255–6264. doi:10.1523/jneurosci.4181-14.2015.

Nemrodov, D., Niemeier, M., Mok, J. N. Y., & Nestor, A. (2016). The time course of individual face recognition: A pattern analysis of ERP signals. *NeuroImage, 132,* 469–476. doi:10.1016/j.neuroimage.2016.03.006.

Nie, A., Griffin, M., Keinath, A., Walsh, M., Dittmann, A., & Reder, L. (2014). ERP profiles for face and word recognition are based on their status in semantic memory not their stimulus category. *Brain Research, 1557*, 66–73. doi:10.1016/j.brainres.2014.02.010.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110(4)*, 611–646. doi:10.1037/0033-295x.110.4.611.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10(9)*, 424–430. doi:10.1016/j.tics.2006.07.005.

Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neuroscience & Biobehavioral Reviews, 34(7)*, 1023–1035. doi:10.1016/j.neubiorev.2009.12.014.

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2011). Complementary Learning Systems. *Cognitive Science, 38(6)*, 1229–1248. doi:10.1111/j.1551-6709.2011.01214.x.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and Neuroscience, 2011*, 1-9. doi:10.1155/2011/156869.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105(32)*, 11087–11092. doi:10.1073/pnas.0805664105.

Osipova, D., Takashima, A., Oostenveld, R., Fernandez, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience, 26(28)*, 7523–7531.

doi:10.1523/jneurosci.1948-06.2006.

Otten, L. J., & Rugg, M. D. (2001). Electrophysiological correlates of memory encoding are task-dependent. *Cognitive Brain Research, 12(1)*, 11–18. doi:10.1016/s0926-6410(01)00015-5.

Otten, L. J., Sveen, J., & Quayle, A. H. (2007). Distinct patterns of neural activity during memory formation of nonwords versus words. *Journal of Cognitive Neuroscience, 19(11)*, 1776–1789. doi:10.1162/jocn.2007.19.11.1776.

Paller, K. A., Voss, J. L., & Boehm, S. G. (2007). Validating neural correlates of familiarity. *Trends in Cognitive Sciences, 11(6),* 243–250. doi:10.1016/j.tics.2007.04.002.

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences, 6(2),* 93–102. doi:10.1016/s1364-6613(00)01845-3.

Parish, G., Hanslmayr, S., & Bowman, H. (2018). The Sync/desync model: how a synchronized hippocampus and a desynchronized neocortex code memories. *The Journal of Neuroscience, 38(14)*, 3428–3440. doi:10.1523/jneurosci.2561-17.2018.

Park, H., Lee, D. S., Kang, E., Kang, H., Hahm, J., Kim, J. S., … Jensen, O. (2014). Blocking of irrelevant memories by posterior alpha activity boosts memory encoding. *Human Brain Mapping, 35(8),* 3972–3987. doi:10.1002/hbm.22452.

Parks, C. M., & Yonelinas, A. P. (2009). Evidence for a memory threshold in second-choice recognition memory responses. *Proceedings of the National Academy of Sciences, 106(28),* 11515–11519. doi:10.1073/pnas.0905505106.

Peirce J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2 (10),* 1-8. doi:10.3389/neuro.11.010.2008.

Pezdek, K., & Blandon-Gitlin, I. (2005). When is an intervening line-up most likely to affect eyewitness identification accuracy? *Legal and Criminological Psychology, 10(2),* 247–263. doi:10.1348/135532505x49846.

Pierce, L. J., Scott, L. S., Boddington, S., Droucker, D., Curran, T., & Tanaka, J. W. (2011). The N250 brain potential to personally familiar and newly learned faces and objects. *Frontiers in Human Neuroscience, 5*. doi:10.3389/fnhum.2011.00111.

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., & Duchaine, B. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Current Biology, 19(4),* 319–324. doi:10.1016/j.cub.2009.01.007.

Pitcher, D., Duchaine, B., Walsh, V., Yovel, G., & Kanwisher, N. (2011). The role of lateral occipital face and object areas in the face inversion effect. *Neuropsychologia, 49(12),* 3448–3453. doi:10.1016/j.neuropsychologia.2011.08.020.

Poch, C., Valdivia, M., Capilla, A., Hinojosa, J. A., & Campo, P. (2018). Suppression of no-longer relevant information in Working Memory: An alpha-power related mechanism? *Biological Psychology, 135,* 112–116. doi:10.1016/j.biopsycho.2018.03.009.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology, 41(2),* 103–146. doi:10.1016/0301-0511(95)05130-9.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60(4),* 437–447. doi:10.1016/j.jml.2009.01.004.

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience, 13(10),* 713–726. doi:10.1038/nrn3338.

Riccio, D. C., Millin, P. M., & Bogart, A. R. (2006). Reconsolidation: A brief history, a retrieval view, and some recent issues. *Learning & Memory, 13(5),* 536–544. doi:10.1101/lm.290706.

Richter, F. R., Cooper, R. A., Bays, P. M., & Simons, J. S. (2016). Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *eLife*, *5*. doi:10.7554/elife.18260.

Roach, B. J., & Mathalon, D. H. (2008). Event-related EEG time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking schizophrenia. *Schizophrenia Bulletin, 34(5),* 907-926. doi:10.1093/schbul/sbn093.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15(1),* 20–27. doi:10.1016/j.tics.2010.09.003.

Roediger, H. L., & DeSoto, K. A. (2013). Confidence and memory: Assessing positive and negative correlations. *Memory, 22(1),* 76–91. doi:10.1080/09658211.2013.795974.

Roediger, H. L., & Karpicke, J. D. (2006). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57(2),* 151–162. doi:10.1016/j.jml.2006.09.004.

Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(5),* 1155–1159. doi:10.1037/0278-7393.31.5.1155.

Rolls, E. T. (2016). Pattern separation, completion, and categorisation in the hippocampus and neocortex. *Neurobiology of Learning and Memory, 129,* 4–28. doi:10.1016/j.nlm.2015.07.008.

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. *Psychological Methods, 22(2),* 304–321. doi:10.1037/met0000057.

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

Rowland, C. A., & DeLosh, E. L. (2014). Mnemonic benefits of retrieval practice at short retention intervals. *Memory, 23(3),* 403–419. doi:10.1080/09658211.2014.889710.

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological Review, 122(1),* 1–23. doi:10.1037/a0037907.

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., Johnson, J. D., & Uncapher, M. R. (2015). Encoding and retrieval in episodic memory. *The Wiley Handbook on the Cognitive Neuroscience of Memory,* 84–107. doi:10.1002/9781118332634.ch5.

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Current Opinion in Neurobiology, 23(2),* 255–260. doi:10.1016/j.conb.2012.11.005.

Rutishauser, U., Aflalo, T., Rosario, E. R., Pouratian, N., & Andersen, R. A. (2018). Single-neuron representation of memory strength and recognition confidence in left human posterior parietal cortex. *Neuron, 97(1),* 209–220.e3. doi:10.1016/j.neuron.2017.11.029.

Schapiro, A. C., McDevitt, E. A., Rogers, T. T., Mednick, S. C., & Norman, K. A. (2018). Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications, 9(1).* doi:10.1038/s41467-018-06213-1.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime (Version 2.0). [Computer software and manual]. Pittsburgh, PA: Psychology Software Tools Inc.

Schwabe, L., Nader, K., & Pruessner, J. C. (2014). Reconsolidation of human memory: brain mechanisms and clinical relevance. *Biological Psychiatry, 76(4),* 274–280. doi:10.1016/j.biopsych.2014.03.008.

Schweinberger, S. R., Huddy, V., & Burton, A. M. (2004). N250r: a face-selective brain response to stimulus repetitions. *NeuroReport, 15(9),* 1501–1505. doi:10.1097/01.wnr.0000131675.00319.42.

Schweinberger, S. R., Pfütze, E.-M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(3),* 722–736. doi:10.1037/0278-7393.21.3.722.

Scully, I. D., Napper, L. E., & Hupbach, A. (2017). Does reactivation trigger episodic memory change? A meta-analysis. *Neurobiology of Learning and Memory, 142,* 99–107. doi:10.1016/j.nlm.2016.12.012.

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review, 18(3),* 455–468. doi:10.3758/s13423-011-0086-9.

Shimamura, A. P. (2011). Episodic retrieval and the cortical binding of relational activity. *Cognitive, Affective, & Behavioral Neuroscience, 11(3),* 277–291. doi:10.3758/s13415-011-0031-4.

Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognitio*n, 39(6), 1712–1725. doi:10.1037/a0033569.

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.

Solomon, C. J., Gibson, S. J., & Mist, J. J. (2013). Interactive evolutionary generation of facial composites for locating suspects in criminal investigations. *Applied Soft Computing, 13(7),* 3298–3306. doi:10.1016/j.asoc.2013.02.010.

St. Jacques, P. L., Olm, C., & Schacter, D. L. (2013). Neural mechanisms of reactivation-induced updating that enhance and distort memory. *Proceedings of the National Academy of Sciences, 110(49)*, 19671–19678. doi:10.1073/pnas.1319630110.

St. Jacques, P. L., & Schacter, D. L. (2013). Modifying memory: selectively enhancing and updating personal memories for a museum tour by reactivating them. *Psychological Science, 24(4),* 537-543. doi:10.1177/0956797612457377.

Staresina, B. P., Alink, A., Kriegeskorte, N., & Henson, R. N. (2013). Awake reactivation predicts memory in humans. *Proceedings of the National Academy of Sciences, 110(52),* 21159–21164. doi:10.1073/pnas.1311989110.

Staresina, B. P., & Wimber, M. (2019). A Neural Chronometry of Memory Recall. *Trends in Cognitive Sciences, 23(12)*, 1071–1085. doi:10.1016/j.tics.2019.09.011.

Staudigl, T., & Hanslmayr, S. (2013). Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. *Current Biology, 23(12),* 1101–1106. doi:10.1016/j.cub.2013.04.074

Steblay, N. K., & Dysart, J. E. (2016). Repeated eyewitness identification procedures with the same suspect. *Journal of Applied Research in Memory and Cognition, 5(3),* 284–289. doi:10.1016/j.jarmac.2016.06.010.

Steblay, N. K., Tix, R. W., & Benson, S. L. (2013). Double exposure: the effects of repeated identification lineups on eyewitness accuracy. *Applied Cognitive Psychology, 27*, 644-654. doi:10.1002/acp.2944.

Sun, S., Fidalgo, C., Barense, M., Lee, A., Cant, J., & Ferber, S. (2017). Erasing and blurring memories: The differential impact of visual interference on separate aspects of forgetting. *Journal of Vision, 17(10)*, 1112. doi:10.1167/17.10.1112.

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology, 15(3),* e2000797. doi:10.1371/journal.pbio.2000797.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences, 3(4)*, 151–162. doi:10.1016/s1364-6613(99)01299-1.

Tanaka, J. W., Curran, T., Porterfield, A. L., & Collins, D. (2006). Activation of preexisting and acquired face representations: the n250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience, 18(9)*, 1488–1497. doi:10.1162/jocn.2006.18.9.1488.

Thakral, P. P., Wang, T. H., & Rugg, M. D. (2015). Cortical reinstatement and the confidence and accuracy of source memory. *NeuroImage, 109*, 118–129. doi:10.1016/j.neuroimage.2015.01.003.

Topp-Manriquez, L. D., McQuiston, D., & Malpass, R. S. (2014). Facial composites and the misinformation effect: How composites distort memory. *Legal and Criminological Psychology, 21(2)*, 372–389. doi:10.1111/lcrp.12054.

Tredoux, C. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology: Applied, 8(3),* 180–

193. doi:10.1037/1076-898x.8.3.180.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97-136. doi:10.1016/0010-0285(80)90005-5.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson, *Organization of memory*. Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie canadienne, 26(1),* 1–12. https://doi.org/10.1037/h0080017.

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

Urquhart, J. A., & O’Connor, A. (2018). Constructing and model-fitting receiver operator characteristics using continuous data. doi:10.31234/osf.io/qdgjc.

Van den Honert, R. N., McCarthy, G., & Johnson, M. K. (2016). Reactivation during encoding supports the later discrimination of similar episodic memories. *Hippocampus, 26(9),* 1168–1178. doi:10.1002/hipo.22598.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 43(2),* 161–204. doi:10.1080/14640749108400966.

Valentine, T. (2001). Face-space models of face recognition. In: M.J. Wenger & J.T. Townsend (eds.) *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 83-113). Mahwah: LEA.

Valentine, T., Davis, J. P., Memon, A., & Roberts, A. (2011). Live showups and their influence on a subsequent video line-up. *Applied Cognitive Psychology, 26(1),* 1–23. doi:10.1002/acp.1796.

Valentine, T., Lewis, M. B., & Hills, P. J. (2015). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology, Jan 27,* 1-24. doi:10.1080/17470218.2014.990392.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(3),* 582–600. doi:10.1037/0278-7393.26.3.582.

Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. *Neuropsychologia, 46(7),* 1787–1799. doi:10.1016/j.neuropsychologia.2008.01.004.

Vilberg, K. L., & Rugg, M. D. (2009). Functional significance of retrieval-related activity in lateral parietal cortex: Evidence from fMRI and ERPs. *Human Brain Mapping, 30(5),* 1490–1501. doi:10.1002/hbm.20618.

Vogelsang, D. A., Gruber, M., Bergström, Z. M., Ranganath, C., & Simons, J. S. (2018). Alpha oscillations during incidental encoding predict subsequent memory for new “foil” information. *Journal of Cognitive Neuroscience, 30(5),* 667–679. doi:10.1162/jocn_a_01234.

Voss, J. L., & Paller, K. A. (2008). Brain substrates of implicit and explicit memory: The importance of concurrently acquired neural signals of both memory types. *Neuropsychologia, 46(13),* 3021–3029. doi:10.1016/j.neuropsychologia.2008.07.010.

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25(1),* 58–76. doi:10.3758/s13423-017-1323-7.

Waldhauser, G. T., Johansson, M., & Hanslmayr, S. (2012). Alpha/beta oscillations indicate inhibition of interfering visual memories. *Journal of Neuroscience, 32(6),* 1953–1961. doi:10.1523/jneurosci.4201-11.2012.

Watanabe, T. (2001). Effects of constrained choice on memory: the extension of the multiple-cue hypothesis to the self-choice effect. *Japanese Psychological Research, 43(2),* 98–103. doi:10.1111/1468-5884.00165

Watanabe, T., & Soraci, S. A. (2004). The self-choice effect from a multiple-cue perspective. *Psychonomic Bulletin & Review, 11(1),* 168–172. doi:10.3758/bf03206478.

Weil, R. S., & Rees, G. (2010). Decoding the neural correlates of consciousness. *Current Opinion in Neurology, 23(6),* 649–655. doi:10.1097/wco.0b013e32834028c7.

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11(6),* 571–580. doi:10.1080/09658210244000414.

Wiese, H., Tüttenberg, S. C., Ingram, B. T., Chan, C. Y. X., Gurbuz, Z., Burton, A. M., & Young, A. W. (2019). A robust neural index of high face familiarity. *Psychological Science, 30(2),* 261–272. doi:10.1177/0956797618813572.

Wilding, E. L., & Ranganath, C. (2011). Electrophysiological correlates of episodic memory processes. In E. S. Kappenman,, & S. J. Luck (Eds.). *The Oxford Handbook of Event-Related Potential Components*. New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780195374148.013.0187.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114(1),* 152–176. doi:10.1037/0033-295x.114.1.152.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117(4),* 1025–1054. doi:10.1037/a0020874.

Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the reliability of eyewitness memory. *Perspectives on Psychological Science, 13(3),* 324–335. doi:10.1177/1745691617734878.

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.

Woroch, B., & Gonsalves, B. D. (2010). Event-related potential correlates of item and source memory strength. *Brain Research, 1317,* 180–191. doi:10.1016/j.brainres.2009.12.074.

Wynn, S. C., Daselaar, S. M., Kessels, R. P. C., & Schutter, D. J. L. G. (2019). The electrophysiology of subjectively perceived memory confidence in relation to recollection and familiarity. *Brain and Cognition, 130,* 20–27. doi:10.1016/j.bandc.2018.07.003.

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences, 34(10),* 515–525. doi:10.1016/j.tins.2011.06.006.

Yates, S. Q. (2017, January 6). *Memorandum for heads of department law enforcement components, all department prosecutors*. Washington, DC: Office of the Deputy Attorney General, U.S. Department of Justice. Retrieved from https://www.justice.gov/archives/opa/press-release/file/923201/download.

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

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: rocs and conscious recollection. *Consciousness and Cognition, 5(4),* 418–441.

doi:10.1006/ccog.1996.0026.

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin, 133(5)*, 800–832. doi:10.1037/0033-2909.133.5.800.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M, D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience, 25(11)*, 3002–3008. doi:10.1523/jneurosci.5295-04.2005.

Appendix

Appendix A. Validating learning procedure used for encoding of face images

The first pilot experiment reported in this thesis was designed in order to ensure that participants were able to encode the faces sufficiently well during the learning phase. That is, recognition tests in the updating experiments were expected to produce a large number of recognition errors because targets were presented amongst distractor faces that, whilst varying to some extent with regards to the internal facial features, were still rather similar to the targets. However, poor recognition performance with these stimuli could be due to target-distractor similarity creating a very difficult task or could be due to insufficient encoding of target faces during learning. Ensuring that faces were adequately encoded is necessary for the experiments in the present thesis to conclude that any biases created during initial recognition testing were due to retrieval-induced *updating* of stored face representations. Therefore, an old/new recognition task was conducted prior to the experiments in chapter 3 to ensure that the learning manipulations used in experiments 3a and 3b were sufficient to facilitate encoding of target faces, by testing whether participants could distinguish these target faces from perceptually dissimilar novel distractor face images during a subsequent recognition task.

A.1 Method

Participants

Twenty-two participants (M_{age} = 19.19, SD_{age} = 0.93), including 21 females with an age range of 18-21, were recruited from University of Kent’s School of Psychology, taking part for undergraduate course credit. All provided informed consent, with ethical approval gained from the School of Psychology ethics committee. All participants had normal/corrected to normal vision and were at least 18 years old. Participants were excluded from this experiment if they were not aged between 18-35, nor if they had participated in any experiment previously within this thesis. None of the participants were recruited for any of the experiments in any of the experimental chapters.

Stimuli and Apparatus

Face stimuli were initially sampled from a face space constructed of 29 components. From this model, sixty face sets (named face sets 1-60) were selected (from a larger batch of 108 face sets that was used for Experiment 3a), each face set containing five faces (images 1-5). Initially, 60 face locations were sampled from the face space origin, with target to target distances ranging from ~4000 - ~7000 (see Figure A.1). For these 60 initial face locations, four additional face locations were randomly sampled from a uniform distribution around the initial face, with a limit of +/- four standard deviations. As demonstrated in Figure A.2, the perceptual dissimilarity between the five face images within each set corresponds to a larger Euclidean distance. One-hundred and twenty faces were selected from this face set batch, with 60 faces defined as ‘old’ faces shown during learning and recognition, and 60 faces used as ‘new’ faces presented only during the recognition test. The 60 face targets were the same chosen for experiments in chapter three, with a single face image taken from face sets 1-60. The 60 ‘new’ faces were taken from sets 61-108, with face images 1-5 selected an equal number of times. These sets were split into two blocks, each block contained 30 ‘old’ and 30 ‘new’ faces, with old faces from sets 1-30 presented with new faces from sets 61-90 (group 1), and old faces from sets 31-60 shown with new faces from sets 91-108 (group 2). Group order was counter-balanced across participants. Note that ‘new’ faces selected for group 2 had to be selected from a limited number of 18 face sets (91-108) due to the face space model constraints that created only 108 face sets in this batch. As the present experiments were used to validate the suitability of images used in Experiment 3a and 3b, it was not possible to create a separate batch of face sets in case there was perceptual overlap between any new face sets and the face images from the original sets. Consequently, it was decided to select 30 faces from face sets 91-108, however it was ensured that all of the ‘new’ faces selected from these sets were distinguishable so as to not have perceptually similar face images in the ‘new’ condition for group 2 (i.e. I would select the two most dissimilar faces from within single face sets).

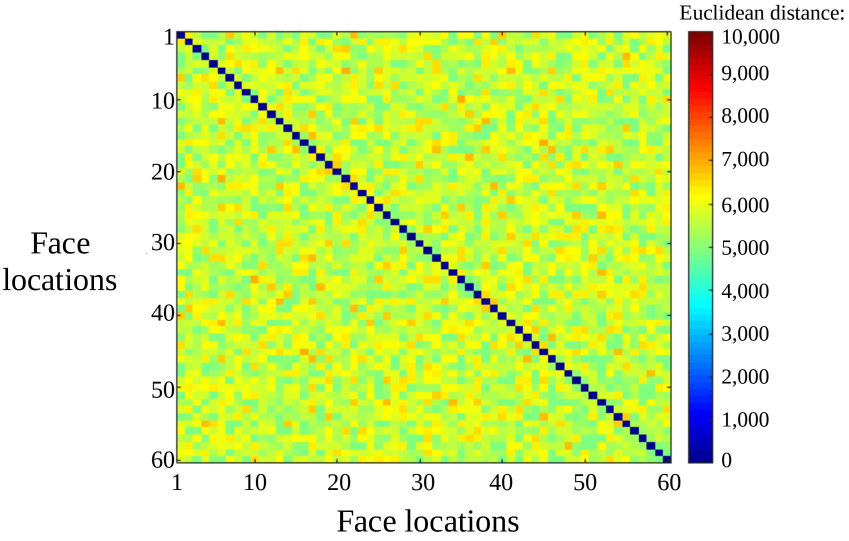

Figure A.1. Euclidean distance between initial face locations. Colour map representing Euclidean distances between the 60 face locations within the image face space from which face stimuli were created.

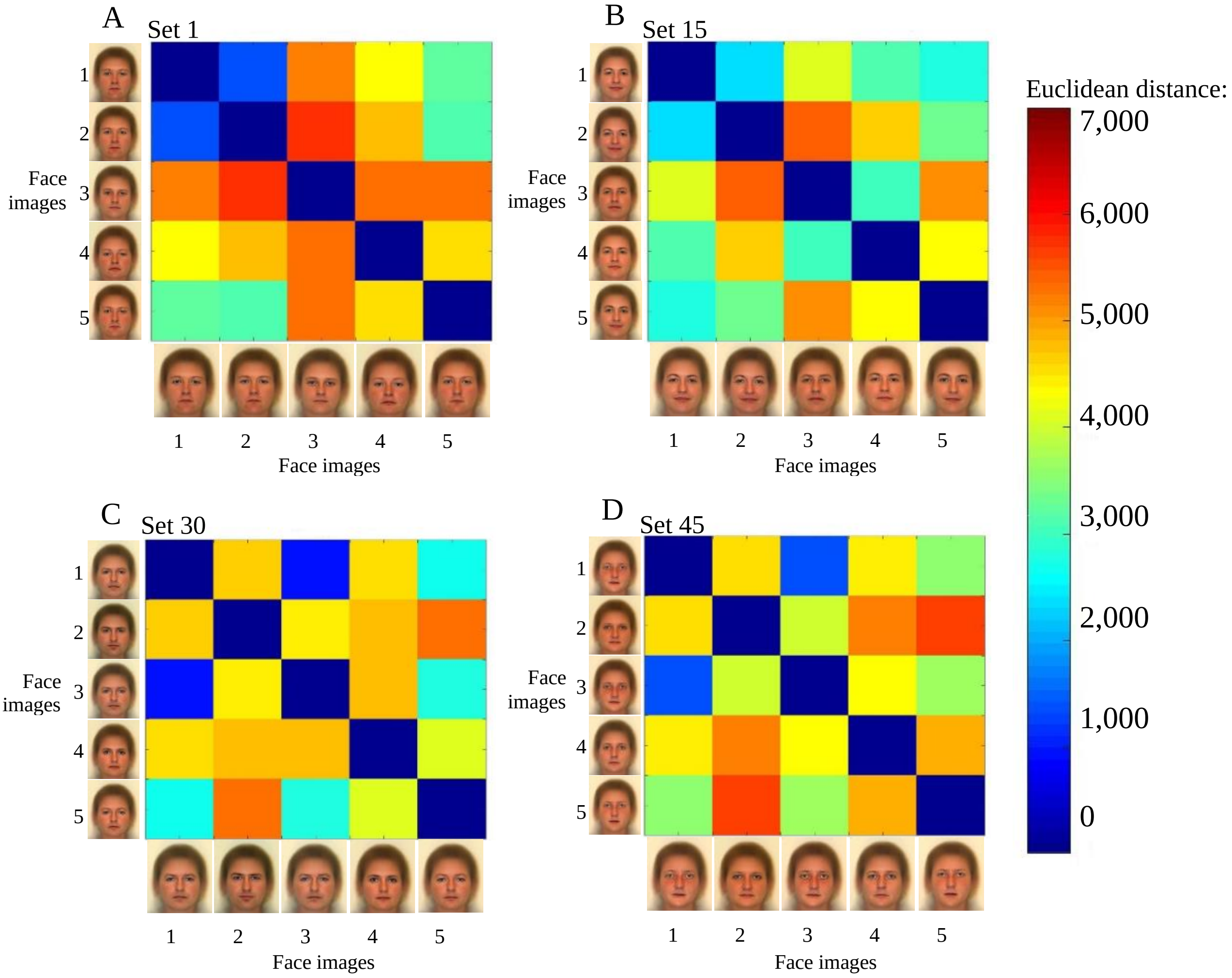

Figure A.2. Euclidean distance between faces within sets. Colour maps representing the Euclidean distances between face images within face set 1 (panel A), face set 15 (panel B), face set 30 (panel C) and face set 45 (panel D). For each face set, perceptual dissimilarity between a pair of faces corresponds to a larger Euclidean distance between face images in the constructed artificial face space.

Procedure and design

Participants completed two blocks, each block consisting of a learning, filler and recognition test procedure (see Figure A.3).

Learning. During learning, participants completed three cycles of face rating tasks, with all 30 face targets shown within each of the three cycles. For each cycle, face targets were presented one at a time, in a randomised order for 5000ms each, preceded by a 500ms fixation. All face pictures subtended a 5.12 x 5.88 visual angle when participants sat at a distance of 75cm from the screen. During face presentation, participants were instructed to focus on the internal features (i.e., eyes, eyebrows, nose, mouth) and were encouraged to memorise the face for a later recognition test. Additionally, participants completed a rating task for each face target to facilitate encoding. For cycle 1, participants rated each face as to whether they thought the person of the face was nice (key press 5) or not nice (key press 1). In cycle 2, participants rated each face as to whether they thought the face was attractive (key press 5) or unattractive (key press 1). Finally, during cycle 3, participants rated each face as to whether they thought the face was female (key press 5) or male (key press 1). Participants were informed of each rating task prior to beginning each cycle and responded whilst the face was on screen.

Letter search filler task. Once the learning task was complete, participants completed a letter search visual task using stimuli from Treisman and Gelade (1980). The purpose of this task was to separate memory encoding and recognition tasks so that performance during the recognition task was based on long-term memory rather than working memory. Seventy-two letter arrays containing a mixture of letter type (X, O & N), frequency (1, 5, 15 & 30) and colour (red, green & blue) combinations were presented one at a time. Participants were tasked with searching for a blue letter, which was present in 18/72 pictures. Participants indicated whether the array had a blue letter (key press 5) or did not have a blue letter (key press 1). Trials were self-paced, with all 72 trials completed in around one minute.

Recognition test. Following the filler task, participants completed a face recognition test. Here, the 30 ‘Old’ faces from learning were randomly intermixed with 30 ‘New’ faces. Each face was shown individually and was on screen until participants made a keyboard response. All face images subtended a visual angle of 5.12 x 5.88. Participants were instructed to decide whether or not the face was shown during learning (Old) or not shown during learning (New). Participants were given response options of 1 = *Definitely new*, 2 = *Possibly new*, 4 = *Possibly old* or 5 = *Definitely old*, allowing a confidence response for Old and New recognition judgements to be made. After making a key press, a 500ms fixation cross was shown before the next face was shown. Once all first block trials were completed, participants had a short rest before completing the second block.

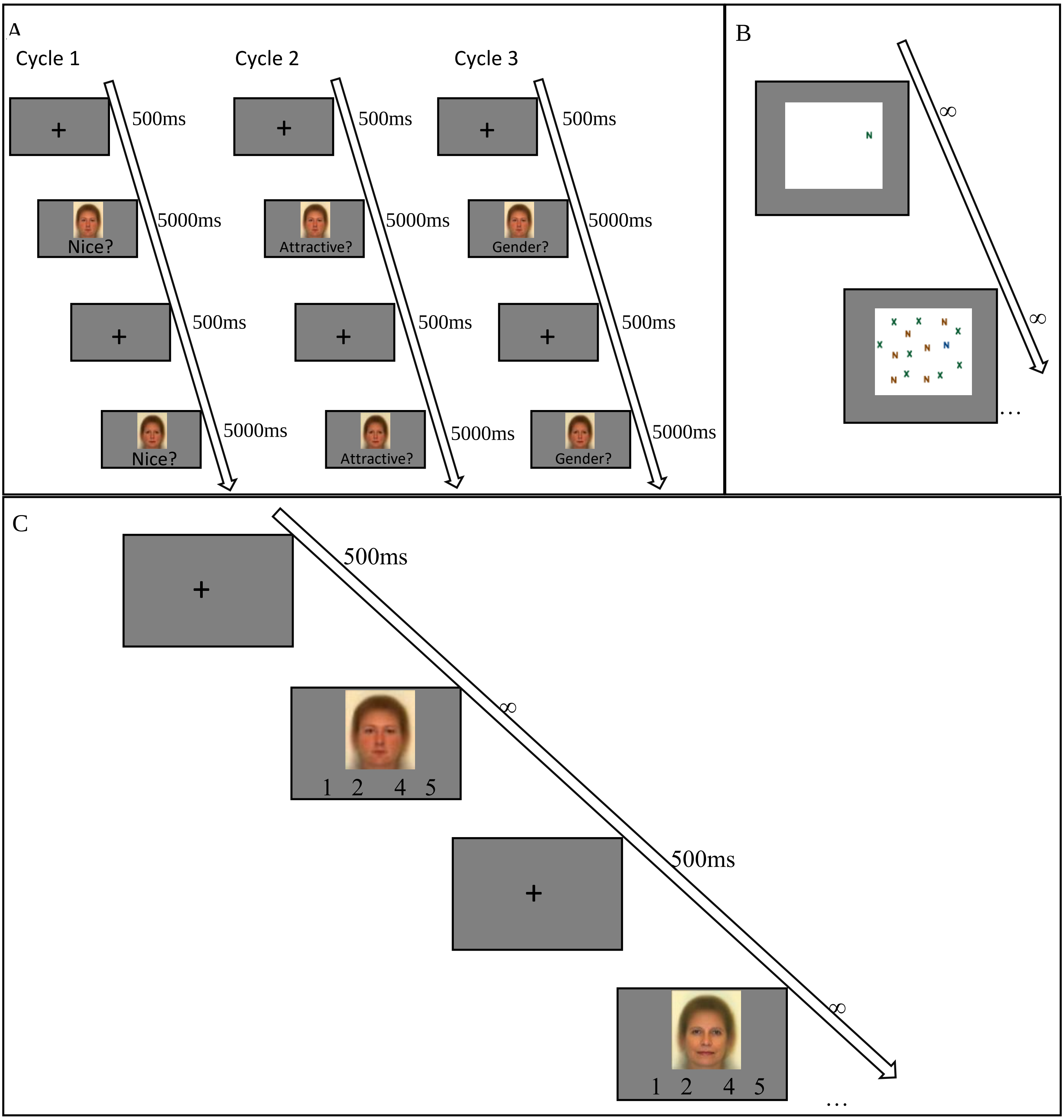

Figure A.3. Procedure for old/new recognition pilot experiment. Panel A, participants learnt thirty face targets in a three cycle procedure. Panel B, participants saw 72 letter arrays and indicated whether a blue letter was present or absent. Panel C, participants completed an old/new recognition test with the presentation of face stimuli that were Old (shown during learning) and New (not shown during learning), with confidence responses acquired.

A.2. Results

One-hundred and twenty responses were acquired for all participants (60 ‘old’ trials and 60 ‘new’ trials). The hit rate was defined as the proportion of Old trials where participants pressed ‘possibly old’ or ‘definitely old’. Similarly, the correct rejection rate was defined as the proportion of New trials where a ‘possibly new’ or ‘definitely new’ response was given. The mean hit rate ($M = .72$, $SE = .02$) and mean correct rejection rate ($M = .65$, $SE = .03$) were both significantly greater than chance performance (0.5; hits: $t(21) = 10.05$, $p < .001$, $g = 6.94$; correct rejections: $t(21) = 4.98$, $p < .001$, $g = 4.56$). Hit and correct rejection rates ranged from 52-92% and 37-83% respectively. Furthermore, signal detection measures were calculated from this data (Snodgrass & Corwin, 1988). A discriminability index (Pr) was calculated as:

$$Pr = \text{Hits} - (1 - \text{Correct Rejection})$$

where a higher score reflects a greater ability to discriminate between Old and New face images. In addition, a response bias (Br) was calculated as:

$$Br = (1 - \text{Correct Rejection}) / (1 - \text{Pr})$$

where values closer to 0 indicate a more conservative response bias (tendency to guess “new”), values closer to 1 indicate a liberal response bias (tendency to guess “old”), and values close to 0.5 indicate no response bias.

Pr and Br measures were collapsed across block for all participants. The mean Pr index across block ($M = .38$, $SE = .03$) was significantly different from zero ($t(21) = 11.73$, $p < .001$, $g = 2.50$), meaning participants could successfully discriminate between old and new items. In addition, the Br value (*M* = 0.55, *SE* = 0.03) was not significantly different to 0.5 (*t*(21) = 1.67, *p* = .11, *g* = 0.36), thus participants, on average, did not adopt a response bias during the recognition test. Consequently, the learning procedure of 30 targets undergoing three rating cycles was used in the experiments in Chapters 3 and 4. The exact same face images used in the current pilot were used as target faces for experiments in Chapter 3, however different target faces were used in subsequent experiments as a separate face set batch was developed to make certain improvements (see Chapter 4 method section).

For the EEG experiment reported in Chapters 5 and 6, a modified version of the learning task was used. In this experiment, participants were asked to learn 15 faces within a shorter block than that reported previously (participants completed five blocks within two separate EEG sessions), in order to maintain attention and performance across the two EEG sessions. Therefore, a second pilot study was conducted prior to data collection for the full EEG experiment to validate the learning procedure used in the EEG experiment. For this pilot experiment, eight participants with an age range of 18-21 (M_{age} = 19.25, SD_{age} = 0.89; seven female) were recruited in line with the University of Kent ethics committee guidelines. The pilot experiment was similar to that described in Figure A.3, except for the learning task length and the number of blocks completed. During the learning task, participants completed two cycles, with 15 face targets shown for 3000ms each, after a 500ms fixation cross. Participants rated faces in cycle one as not trustworthy (key press 1) or trustworthy (key press 5). In cycle two, participants rated faces as not attractive (key press 1) or attractive (key press 5). During the subsequent old/new recognition task, participants were shown the 15 target faces along with 15 new faces. In total, participants completed five blocks.

As in the previous analysis, a Pr and Br measure was calculated from the hit (*M* =

0.71, *SD* = 0.15) and correct rejection rates (*M* = 0.70, *SD* = 0.08) for each participant. Hit and correct rejection rates ranged from 42-88% and 53-83% respectively. From these eight participants, the mean Pr index (*M* = 0.41, *SE* = .07) was significantly different from zero (*t*(7) = 6.13, *p* < .001, *g* = 2.17). In addition, the mean Br index (*M* = 0.52, *SE* = .04) was not significantly different to a point value of 0.5 (*t*(7) = 0.54, *p* = .61. *g* = 0.19), suggesting the test did not lead to consistent response biases. Although these results should be viewed with caution due to the limited sample size, the mean values alone, given their similarity to the previous results, were used to determine that a two cycle learning procedure was sufficient for participants to learn 15 faces within a single block for the EEG study in Chapter 5.

Appendix B. Validating the relationship between perceptual dissimilarity and Euclidean distance

The face stimuli used in the reported experiments were created from an artificial face image space. As mentioned, a critical assumption of this method is that a pair of faces becomes more perceptually dissimilar the further away they are from one another within image space, as in the theoretical literature on face space in cognition (Valentine et al., 2015). The experiments within the present thesis used a Euclidean distance metric as a continuous measure of face recognition accuracy and updating. In order to verify that larger Euclidean distances between a pair of face images corresponds to a greater perceptual dissimilarity between these image pairs, two validation tasks were conducted, a face similarity task and a face ranking task. These validation tasks were conducted following the completion of Experiments 3a and 3b in order to verify the suitability of face set parameters. As will be demonstrated, information from these validation tasks was used in the design of face stimuli for the experiments in Chapters 4, 5 and 6.

B.1. Method

Participants

Twenty-six undergraduate participants ($M_{age} = 20.19$, $SD_{age} = 3.33$; 21 females) completed the face validation experiment. All participants were naïve to the study aims and face stimuli. Participants received course credits and ethical approval was received from the University of Kent ethics board.

Apparatus

The experiment was completed on a Dell optiplex 9020 desktop computer with E-Prime (Schneider, Eschman, & Zuccolotto, 2002). The screen measured 51cm x 28.4cm, with a resolution of 1920 x 1080.

Procedure

Face similarity task. Participants were presented with 54 trials (sets 1-54 from the batch of 108 face sets, where sets 1-60 were used for the experiments in Chapter 3 and the learning validation experiment in Appendix A), with one face set presented per trial. For each trial, two faces were randomly selected from the five images within each face set and presented horizontally across the screen. Image pairs were presented with a 6-point scale, ranging from 1 ("*Not at all similar*") to 6 ("*The same image"*; see Figure B.1). For each trial, participants rated the perceptual similarity of face pairs using the scale provided. No time limit was imposed, however participants were encouraged to respond as quickly as possible. Following a key press, a fixation cross was presented in the centre of the screen for 500ms before the next trial began. Trial order was randomised for each participant. All trials were completed within 10 minutes.

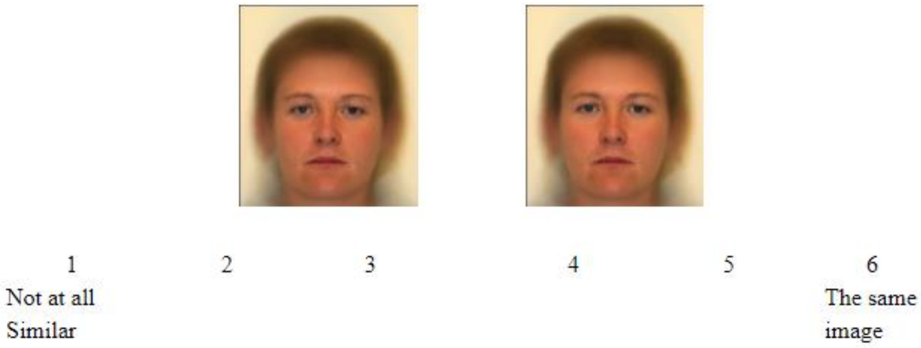

Figure B.1. Example trial for face rating task. Participants were presented with two randomly selected images taken from one face set.

Face ranking task. Following a short break, the remaining 54 face sets (sets 55-108) were used for the face ranking task. For each trial, all five images from the face set were presented simultaneously on screen. One of the five images was designated as a ‘target image’, with the other four images being ‘response faces’, labelled 1, 2, 3 and 4 (see Figure B.2). Images 1, 2, 3, 4 and 5 within a face set were selected as the target face an equal number of times as possible.

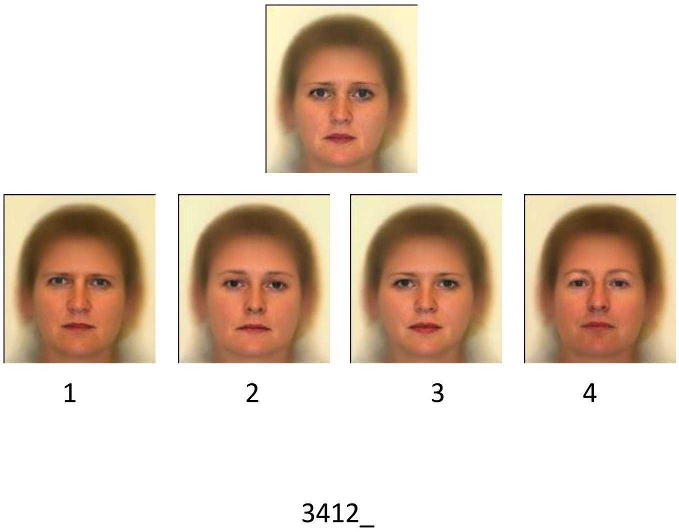

Figure B.2. Example of a face ranking task trial. Participants were asked to rank the response faces in order of similarity to the target face.

Participants were asked to rank the response faces in order of similarity to the target face. In the example in Figure B.2, response faces were ranked by pressing 3412, indicating that image 3 was most similar, image 4 was second similar, image 1 was third similar and image 2 was least similar to the target for this particular face set. Trials were self-paced, and participants could amend their choices before beginning the next trial by pressing the backspace key. Participants received visual feedback with regards to the rank order they had chosen, and were instructed to press the spacebar key to move onto the next trial. A 500ms fixation cross was presented prior to each trial, with trial order randomised for each participant.

B.2. Results

Face rating results

For each of the 54 trials, the similarity rating (1-6) and Euclidean distance between presented faces was extracted using Matlab. Spearman correlations between face ratings and Euclidean distances were then calculated within-participant. The mean Spearman correlation across participants was -.56 (*SE* = 0.02), significantly different from 0 (*t*(25) = 26.23, *p* <.001, *g* = 5.25). Figure B.3 shows that each participant demonstrated a negative relationship between Euclidean distance and face similarity ratings. These results suggest that participants rated face pairs with a smaller Euclidean distance as more perceptually similar, consistent the assumption that Euclidean distance can be used as an accurate, continuous metric of face recognition error.

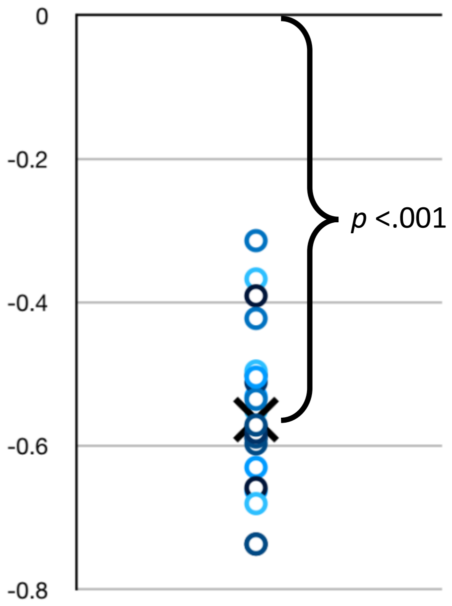

Figure B.3. Relationship between Euclidean distance and perceptual similarity ratings. Scatterplot of mean correlation coefficients at participant (blue circles) and group-level (black cross) representing the relationship between Euclidean distance of face images and face similarity ratings.

Non-linear relationship between Euclidean distance and perpetual similarity. An interesting observation from individual scatterplots of Euclidean distance and face similarity ratings suggested that, for some participants, the negative correlation between Euclidean distance and face similarity ratings were non-linear. For example, Figure B.4 shows the scatterplot for one participant with a correlation coefficient value of -.58. However, inspecting this participant's scatterplot suggests that the negative relationship between Euclidean distance and face similarity was driven by face pairs that had a Euclidean distance between 0 to ~4.0. However, beyond distances of ~4.0, the relationship becomes non-linear.

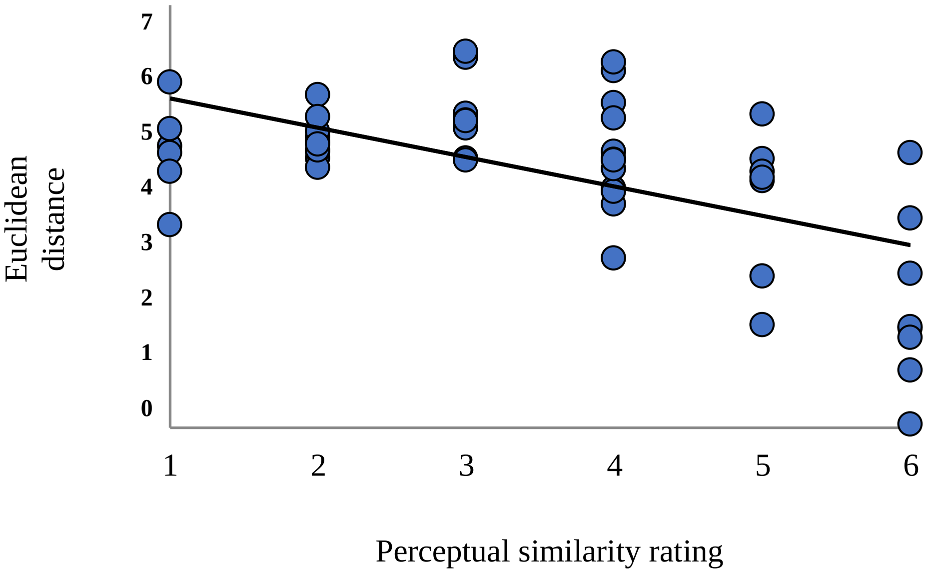

Figure B.4. Correlation between Euclidean distance and perceptual similarity ratings for one participant. Scatterplot suggests a negative but non-linear relationship between Euclidean distance and perceptual similarity of face pairs.

To test the statistical reliability of this observation, Spearman correlations were re-calculated from the face rating data with various Euclidean distance cut-off values. That is, correlations were calculated separately for trials where the Euclidean distance for face pairs was below 3.0 or above 3.0, with mean correlation values computed for these two conditions. The same procedure was done with cut-off values of 3.5, 4.0, 4.5 and 5.0, in order to assess at which distance does the relationship between face similarity and perceptual ratings become non-linear. Table B.1 presents the mean correlation coefficients for these 10 conditions, along with the mean trial number per condition. The mean values for each condition were all negative, with each condition being significantly different from zero (all *p*’s <.001). However, it appears that the relationship between Euclidean distance and face similarity rating starts to weaken when the Euclidean distance between a face pair is more than 3.5. Indeed, paired samples t-tests confirmed that only the 3.0 cut-off condition had a non-significant difference between below and above conditions (*t*(22) = 1.69, *p* = .11, *g* = 0.35). In contrast, the t-tests for the other four cut-off conditions were significant (3.5 – *t*(24) = 4.41, *p* <.001, *g* = 0.88; 4.5 – *t*(24) = 6.78, *p* <.001, *g* = 1.36; 4.5 – *t*(24) = 7.46, *p* <.001, *g* = 1.49; 5.0 – *t*(24) = 6.70, *p* <.001, *g* = 1.34). These results indicate that participant's perception of face pair similarity best corresponds to Euclidean distances from values of 0 to ~3.5. Beyond this value, participants are less likely to identify a face with a Euclidean distance of 4.5 as being more similar to a target face than a face with a Euclidean distance of 5.0 from a target, suggesting the importance of having a varied selection of face images within a face set to optimise the Euclidean distance as a continuous measure of recognition processing.

Table B.1. Relationship between Euclidean distance and perceptual similarity ratings. Mean correlation coefficients and trial numbers for face pairs where the Euclidean distance lay below or above cut-off Euclidean values of 3.0, 3.5, 4.0, 4.5 and 5.0.

Condition	Less than 3.0	More than 3.0	Less than 3.5	More than 3.5	Less than 4.0	More than 4.0	Less than 4.5	More than 4.5	Less than 5.0	More than 5.0
Mean	-0.41	-0.31	-0.50	-0.27	-0.54	-0.19	-0.55	-0.15	-0.56	-0.19
SE	0.05	0.03	0.05	0.03	0.04	0.03	0.03	0.03	0.03	0.04
Trial N	8	46	11	43	17	37	28	26	40	14

Face ranking results

For each participant, Euclidean distance values were extracted between the target face and the response faces selected as the 1st, 2nd, 3rd and 4th rank to the target, for all 54 face sets. Within participants, the mean Euclidean distance values were calculated for each of the four rank categories. Averaged across participants, Figure B.5 demonstrates a monotonic increase in average Euclidean distance between face targets and response faces ranked as 1st, 2nd, 3rd and 4th. A one-way repeated measures ANOVA confirmed a significant main effect of face rank (*F*(2.19,52.62) = 178.85, *p* <.001, η^2p = .88). Furthermore, each pairwise comparison between the four rank categories was statistically significant at *p* <.001. These results indicate that participants could successfully rank a set of four faces in order of similarity to a given target face for the face stimuli used in the current experiments.

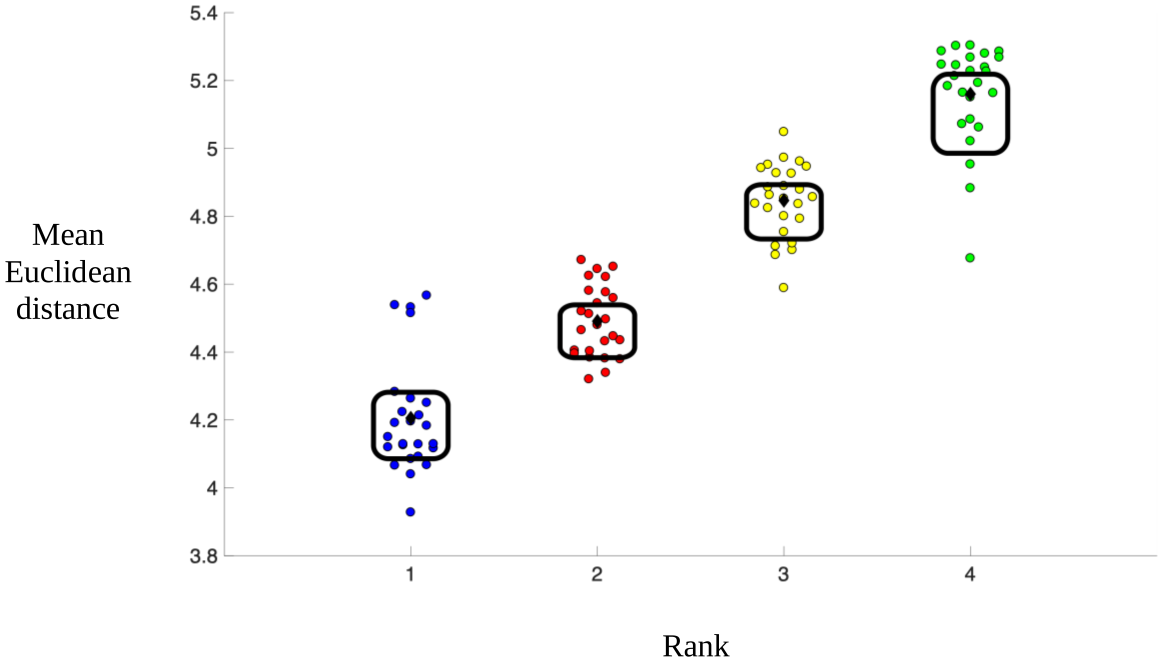

Figure B.5. Euclidean distance for faces ranked 1-4. Participant-level Euclidean distance values for faces that participants ranked as 1st, 2nd, 3rd and 4th closest to targets. Scatterplots represent participants average Euclidean distance between targets and faces selected in each rank category, across all 54 trials. Mean Euclidean distance values were significantly different for all pairwise rank comparisons at p <.001.

Ranking faces within individual face sets. Additional analyses from the face ranking data were conducted to assess participant's ability to rank order individual face sets. The face sets used in the pilot experiments varied in the extent to which the faces within a set were dissimilar from one another. As mentioned, individual face sets were created with limits imposed as to how much distractor faces could vary from the initial target location. However, within these limits, the selection of distractor images was random along a uniform distribution. This random selection process is emphasised when comparing the pairwise Euclidean distances across different face sets. For example, Figure B.6 shows face set 83 (panel A) showing a range of face dissimilarity (Euclidean distance ranged from 0.96 to 4.48) and face set 74 (panel B) showing a limited range of face dissimilarity (Euclidean distance ranged from 3.97 to 4.45). In relation to the face ranking experiment, participants may have found the ranking task simpler when presented with face set 83, but more difficult to rank faces in order of similarity for face set 73, questioning the validity of using sets such as set 73. Therefore, the following analyses were done to identify characteristics of face sets that participants could successfully rank order.

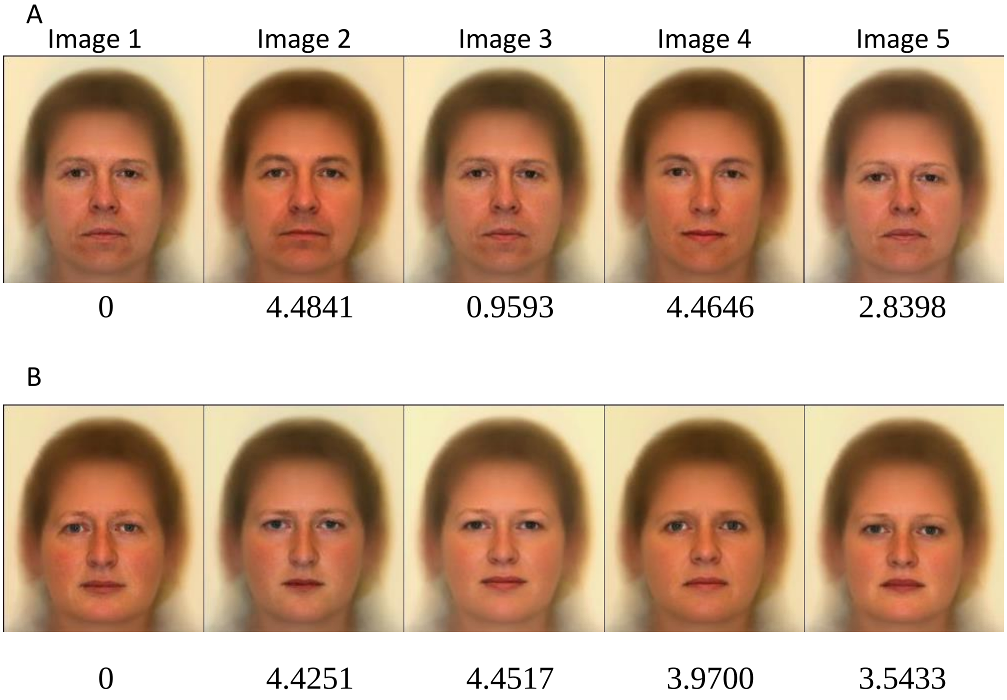

Figure B.6. Example of face image variability within sets. Face sets 83 (panel A) and 74 (panel B), demonstrating how individual face sets from the face generation methods used for the present experiments can produce varied dissimilarity within face sets.

Firstly, the 54 face sets were categorised as ‘Varied’ face sets (N = 33) or ‘Limited’ face sets (N = 21), with a Varied face set defined as a set with Euclidean distance values that spanned a minimum of three boundary Euclidean distances (e.g. panel A in Figure B.6

has faces in three Euclidean boundaries of 0-1 [image 3], 2-3 [image 5] and 4-5 [images 2 & 4]), whereas a Limited face set was defined as a set with Euclidean distance values that spanned a two boundary Euclidean distances or less (e.g. panel B in Figure B.6). For individual Varied and Limited face sets, the Euclidean distance was acquired between target and response images in each rank category. The Euclidean distances for rank 1, rank 2, rank 3 and rank 4 were averaged across participant, for each set, with Figure B.7 showing the distribution of Euclidean distances for ranks 1-4 in Varied and Limited face sets. This figure demonstrates that, for Varied face sets, participants show a monotonic increase in Euclidean distances between target and response faces ranked 1-4. In stark contrast, for Limited face sets, the mean Euclidean distances between target and response faces ranked 1-4 was similar for each rank condition.

To confirm the statistical reliability of these mean values, a 2 (set quality; Varied, Limited) x 4 (image rank; one, two, three, four) repeated measures ANOVA was conducted, showing a significant main effect of set quality ($F(1,20) = 8.84$, $p = .008$, $\eta^2p = 0.31$), whereby Varied face sets had smaller Euclidean distances than Limited face sets, collapsed across rank conditions. In addition, a main effect of rank order was found ($F(1.68,33.56) = 22.89$, $p <.001$, $\eta^2p = 0.53$). As mentioned in the previous analysis, the mean Euclidean distance between rank conditions and target faces followed a monotonic increase from rank one to four, collapsed across set quality. However, a significant interaction also emerged between set quality and rank order ($F(1.63,32.60) = 13.89$, $p <.001$, $\eta^2p = 0.41$). Simple main effects analysis established that average Euclidean distances were significantly smaller for Varied compared to Limited face sets for rank one ($F(1) = 18.98$, $p <.001$), rank two ($F(1) = 9.11$, $p = .007$) and rank three conditions ($F(1) = 6.10$, $p = .02$). However, no significant difference was found between Varied and Limited face sets for the rank four condition ($F(1) = 0.10$, $p = .76$). These findings add more evidence to support the view that face images within sets can only be distinguished with sufficiently varied Euclidean image dissimilarity between faces within a set.

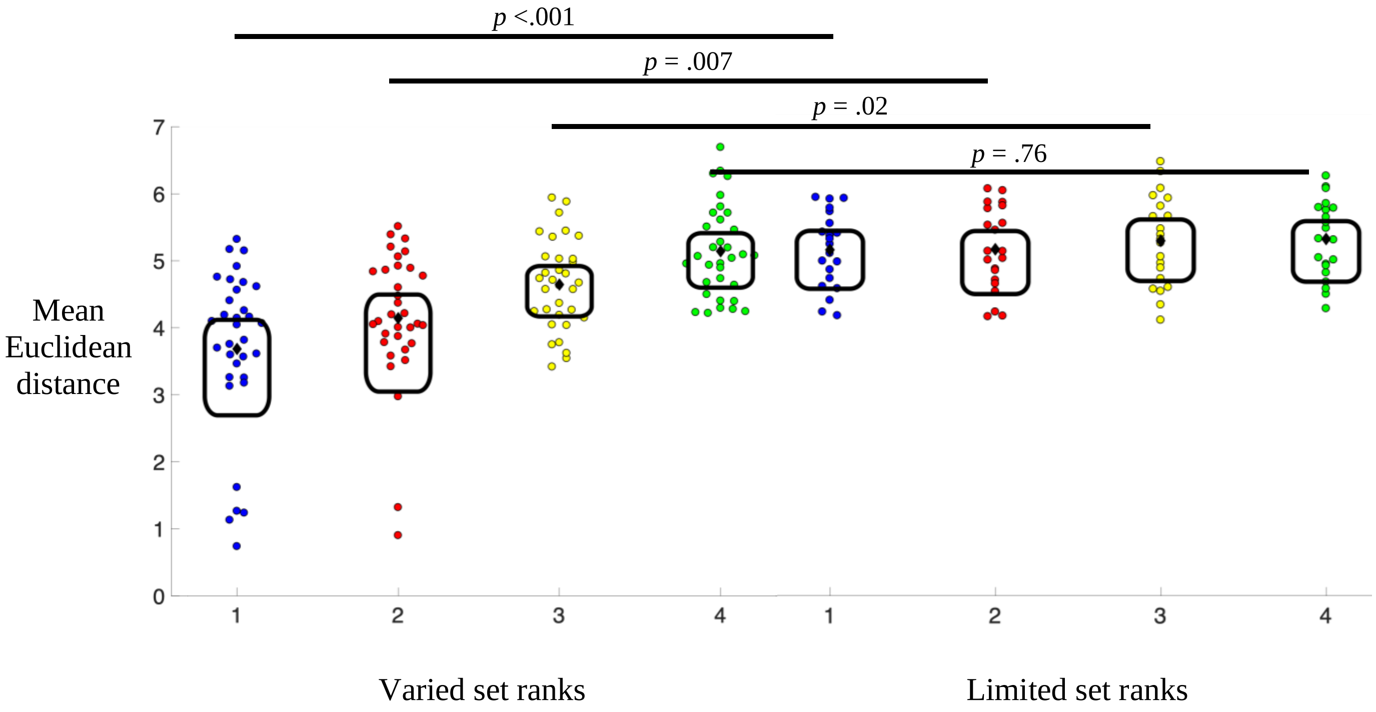

Figure B.7. Euclidean distance for faces ranked 1-4 in sets with varied or limited variability. Scatterplots and mean Euclidean distance values for faces ranked as 1st, 2nd, 3rd and 4th closest to targets for varied in limited sets.

In summary, these findings imply the importance of face image variability within a face set for participants to successfully rank order face images to a given target. This importance is emphasised when applied to the recognition experiments used in the present thesis. Consider, for example, if participants were presented with face set 83 (a variable set) during a recognition trial. Participants may subjectively recognize three of the five images (image 1, 3 and 5), owing to the fact that face images 3 and 5 are physically and perceptually similar to the target face. For this trial, it would therefore be more challenging for participants to correctly select the target face, since doing so may require a very precise memory representation, and/or the use of pattern separation processes (Rolls, 2016; Yassa & Stark, 2011), to aid accurate retrieval. Subsequently, the advantage of having a continuous measure of recognition memory is crucial when analysing sets such as set 83. In contrast, should participants be presented with face set 74 during a recognition trial, the task of selecting the target face becomes much easier due to the fact that none of the four distractor images strongly resemble the target image.

The above considerations were used to adapt the creation of face stimuli for Chapters 4, 5 and 6. Specifically, face sets in these chapters were created such that nine face images were made within a face set. From these nine face images, five images were selected for presentation in the experiment. For each set, image 1 was selected as this image was the original face location when sampling images from face space. The remaining four images were chosen so that, within a set, a range of Euclidean distances between image 1 and the remaining four images were present and when possible, covered as many Euclidean boundaries as possible. Selection of face sets was verified between two researchers to ensure a reliable selection of face sets that contained sufficient perceptual variability, ensuring that multiple Euclidean distance boundaries (i.e., 1000, 2000, 3000, 4000+) were included within a single face set.

Appendix C. Supplementary results for Experiment 3a

Appendix C.1. Proportion confidence across Test 1 and 2.

In Experiment 3a, the analysis of proportion accuracy, confidence and reason responses for Test 1, Test 2 repeated and Test 2 baseline conditions demonstrated that objective accuracy and subjective retrieval experience did not covary across conditions in a simple way (see Section 3.1.2). Specifically, despite similar accuracy for the repeated and baseline condition during Test 2, participants reported more ‘multiple memory’ reasons for giving low confidence responses for the repeated condition during Test 2. This finding emphasises how repeated retrieval may lead to a competition between the target face and faces selected during Test 1 during subsequent Test 2 recognition. Furthermore, these findings suggest that the numerical tendency towards reduced recognition accuracy from Test 1 to Test 2 (although not statistically significant) resulted from different causes depending on the trial types; interference from distractor faces encoded during Test 1 for repeated trials, and delay between encoding and retrieval for baseline trials.

Whilst recognition accuracy at Test 1 was higher than Test 2 accuracy for the repeated condition, participant confidence for these conditions was not significantly different. Building on this finding, confidence measures for repeated trials were analysed for different types of memory responses across Test 1 and 2, to examine whether certain combinations of recognition responses increased confidence during Test 2. For this, repeated trials during Test 2 were initially categorised by the accuracy response on the preceding Test 1 trial (Test 1 correct, Test 1 incorrect). Test 1 correct trials were then split according to accuracy during Test 2 recognition, leaving trial categories of Test 1 correct-Test 2 correct and Test 1 correct-Test 2 incorrect. In addition, for Test 1 incorrect trials, trials were split according to whether these incorrect responses were re-selected during Test 2 (i.e., recognition bias). That is, trials were categorised by whether the same incorrect face selected during Test 1 was chosen during Test 2 recognition (Test 1 incorrect-Test 2 bias), or whether an incorrect face was chosen during Test 1 but not re-selected during Test 2 (Test 1 incorrect-Test 2 non-bias). For each of these four categories, proportion of high confidence responses during Test 2 was calculated.

Table C.1 demonstrates that confidence was greatest for Test 1 correct-Test 2 correct trials, which was significantly superior to confidence for Test 1 correct-Test 2 incorrect ($t(37) = 4.01$, $p < .001$, $g = 0.84$), Test 1 incorrect-Test 2 bias ($t(37) = 3.07$, $p = .004$, $g = 0.64$) and Test 1 incorrect-Test 2 non-bias trials ($t(37) = 3.73$, $p = .001$, $g = 0.71$). No significant effects appeared between the other three measures (all $p > .10$). Therefore, elevated confidence during Test 2 recognition was predominantly driven by trials where the correct face was recognised across multiple recognition attempts.

Table C.1. Descriptive statistics of proportion high confidence responses as a function of Test 1 accuracy and Test 2 bias.

Proportion high confidence	Test 1 correct Test 2 correct	Test 1 correct Test 2 incorrect	Test 1 incorrect Test 2 bias	Test 1 incorrect Test 2 non-bias
Mean	0.67	0.49	0.50	0.50
Standard Error	0.04	0.03	0.04	0.04

Appendix C.2. Relationship between accuracy and confidence.

The analysis in Experiment 3a (Section 3.1.2) examined proportion measures of objective and subjective retrieval during face recognition. Following this, a supplementary analysis examined how the relationship between recognition accuracy and confidence changed as a function of initial, repeated and delayed testing. These analyses built on the advantage of having Euclidean distance as a continuous metric of recognition accuracy (or error, in this case, as larger values indicate a greater distance between recognised and target faces). This allowed for the analysis of how confidence correlated to objective retrieval decisions of all recognition trials, rather than just those where target faces were selected. For trials within each condition, the Euclidean distance between faces selected during recognition and the target face was acquired (Euclidean error). In addition, the confidence response for each trial was acquired. Next, Spearman's correlations were conducted to assess the relationship between Euclidean error and recognition confidence, separately for Test 1, Test 2 repeated and Test 2 baseline conditions. For Test 1 trials, the mean Spearman coefficient was not significantly different to zero ($t(37) = 1.36$, $p = .18$, $g = 0.22$), suggesting no consistent relationship between recognition accuracy and confidence during Test 1. However, during Test 2, Spearman's correlations were significantly less than zero, for both repeated ($t(37) = 5.51$, $p <.001$, $g = 0.89$) and baseline conditions ($t(37) = 2.98$, $p = .005$, $g = 0.48$). The negative relationships between Euclidean error and confidence (see Figure C.1) during Test 2 shows that higher confidence was related to smaller Euclidean error values, specifically, increased confidence was associated with more accurate recognition responses.

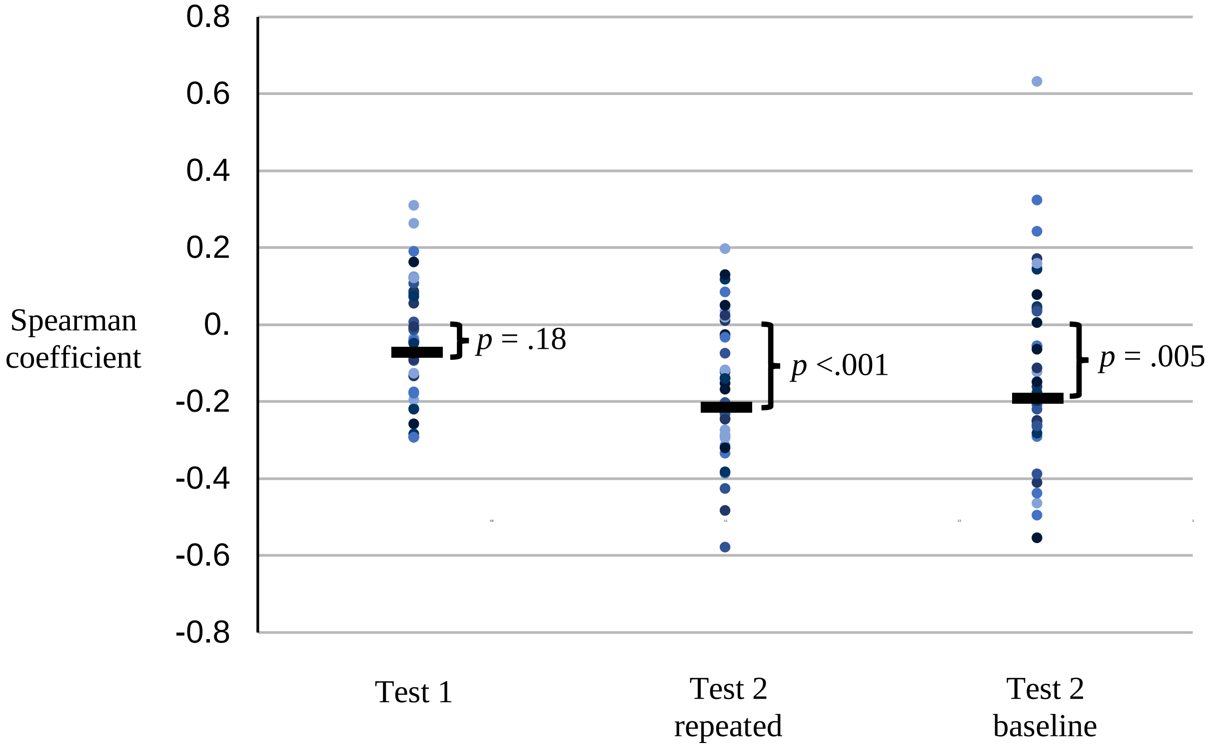

Figure C.1. Accuracy and confidence correlations during Test 1 and 2. Participant-level Spearman coefficients for the relationship between recognition error and confidence during Test 1, Test 2 repeated and Test 2 baseline conditions. Mean coefficients (black horizontal bars) were significantly different from zero for Test 2 repeated and baseline conditions, but not Test 1.

Appendix C.3. Supplementary analysis of Test 2 recognition bias using bias difference measure.

In Experiment 3a, the key measure of face memory updating was the proportion of Test 2 trials where the same face was selected in the previous Test 1 task. For "prior error" trials (i.e. where a distractor was selected during Test 1), an alternative measure of updating was also calculated, the final recognition bias difference, which corrected proportion of incorrect bias responses by the proportion of incorrect non-bias responses (whereby a distractor selected during Test 2 was different to the distractor selected during Test 1). The bias difference measure was significantly greater than zero for the incorrect high condition (*M* = 0.19, *SE* = 0.04; *t*(36) = 5.37, *p* <.001), showing that participants were more likely to select the same face between Test 1 and 2, rather than switch recognition responses during Test 2, for these three conditions. However, mean bias difference was not different to zero for the incorrect low condition (*M* = 0.05, *SE* = 0.03; *t*(36) = 1.99, *p* = .05), meaning that participants were just as likely to select a biased or non-biased distractor during Test 2. A paired samples t-tests confirmed that the bias difference measure was significantly greater following high versus low confidence for incorrect Test 1 conditions (*t*(36) = 4.07, *p* <.001, *g* = 0.66). These results lead to the same conclusion drawn from analysing incorrect bias without the correction of non-bias responses, thus validating this approach as a measure of recognition bias and memory updating for Chapter 4 (see section 4.1).

Appendix D. Supplementary results for Experiment 3b

Appendix D.1. Relationship between accuracy and confidence.

For Test 1, Test 2 repeated and Test 2 baseline conditions across the experiment, the relationship between accuracy and confidence was calculated to again assess how repeated retrieval affected the relationship between objective and subjective recognition, utilising continuous measures of recognition accuracy so that all trials within each condition were analysed. Within conditions, the Euclidean distance between faces selected during recognition and the target face for each trial was calculated (recognition error). Similar, the keypress duration for each trial was acquired as a measure of recognition confidence. Pearson's correlations were conducted on Euclidean error and confidence responses, separately for Test 1, repeated and baseline trials. The mean correlation coefficients (see Figure D.1) for all three conditions were significantly different from zero (Test 1 - *t*(53) = 5.73, *p* <.001, *g* = 0.78; Test 2 repeated - *t*(53) = 6.00, *p* <.001, *g* 0.82; Test 2 baseline - *t*(53) = 3.39, *p* = .001, *g* = 0.46), with Bayes Factors providing more evidence in favour of the hypothesis that the mean coefficient would be more negative than zero (Test 1 – $\mathrm{BF}_{-0(0.22,0.707)}$ = 58235.61; Test 2 repeated – $\mathrm{BF}_{-0(0.89,0.707)}$ = 125486.35; Test 2 baseline – $\mathrm{BF}_{-0(0.48,0.707)}$ = 36.33). The consistent negative correlations between recognition error and confidence for all conditions shows that participants had increased confidence with more accurate recognition responses. These results are generally similar to those seen in Experiment 3a, particularly when comparing the descriptive statistics between experiments. The only difference between experiments was that the Test 1 accuracy-confidence correlation was not significant in Experiment 3a.

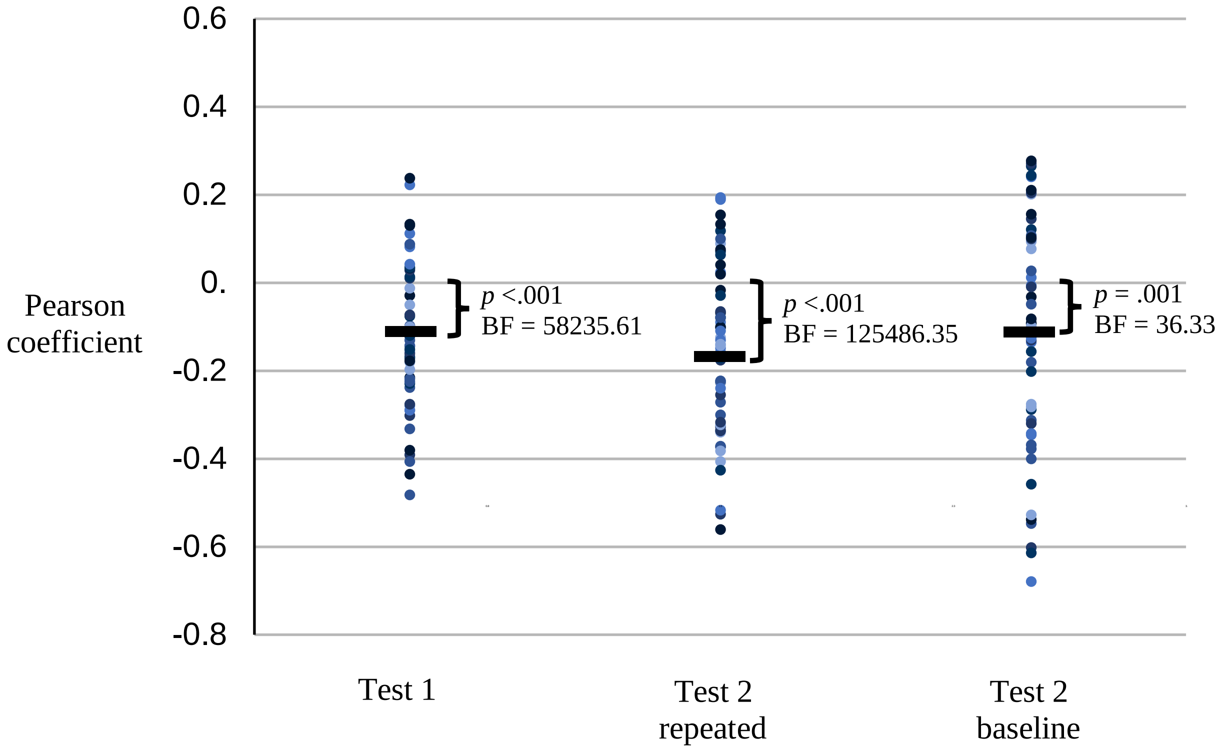

Figure D.1. Correlation between accuracy and confidence during Test 1 and 2. Scatterplots of Pearson coefficients assessing the relationship between Euclidean accuracy and recognition confidence during Test 1, Test 2 repeated and Test 2 baseline conditions. Mean coefficients (black horizontal bars) were significantly below zero for all three conditions.

Appendix D.2. Supplementary analysis of Test 2 recognition bias using bias difference measure.

Similar to Experiment 3b, a bias difference measure was calculated for "prior error" trials, compared between "prior error" trials made with high vs. low confidence. The mean bias difference significantly higher than a point value of zero for Test 1 incorrect high confidence (M = 0.13, SE = 0.03; $t(50) = 4.84$, $p <.001$, g = 0.68), however the mean bias difference measure for the incorrect low confidence condition was not significantly different from zero (M = 0.04, SE = 0.02; $t(39) = 1.84$, $p = .07$, g = 0.26). Planned comparisons showed that the bias difference measure was larger for the high compared to low confidence condition (*t*(50) = 2.24, *p* = .03, *g* = 0.31), though the Bayes factor for the incorrect high-low confidence contrast provided only anecdotal evidence in favour of the alternative versus null hypothesis ($BF_{+0(0,0.67)}$ = 1.87). The similarity of results between proportion bias and the bias difference measure again verify the use of the bias difference measure in the experiments of Chapter 4.

Appendix D.3. Confirming validity of median split confidence analysis in Experiment 3b.

The analysis of recognition bias in Experiment 3b required the creation of artificial high and low confidence conditions, based on the median split of keypress lengths for each response during Test 1. To confirm the validity of the median split analysis, mean keypress length (seconds) was calculated within each of the Test 1 correct high, correct low, incorrect high and incorrect low confidence conditions. Figure D.2 demonstrates how all 54 participants displayed higher mean keypress lengths for the high confidence compared to low confidence conditions. Consequently, the robustness of the findings from Experiment 3b, that recognition confidence modulates retrieval-induced updating, was confirmed by the success of the median split manipulation to create artificial high and low confidence conditions.

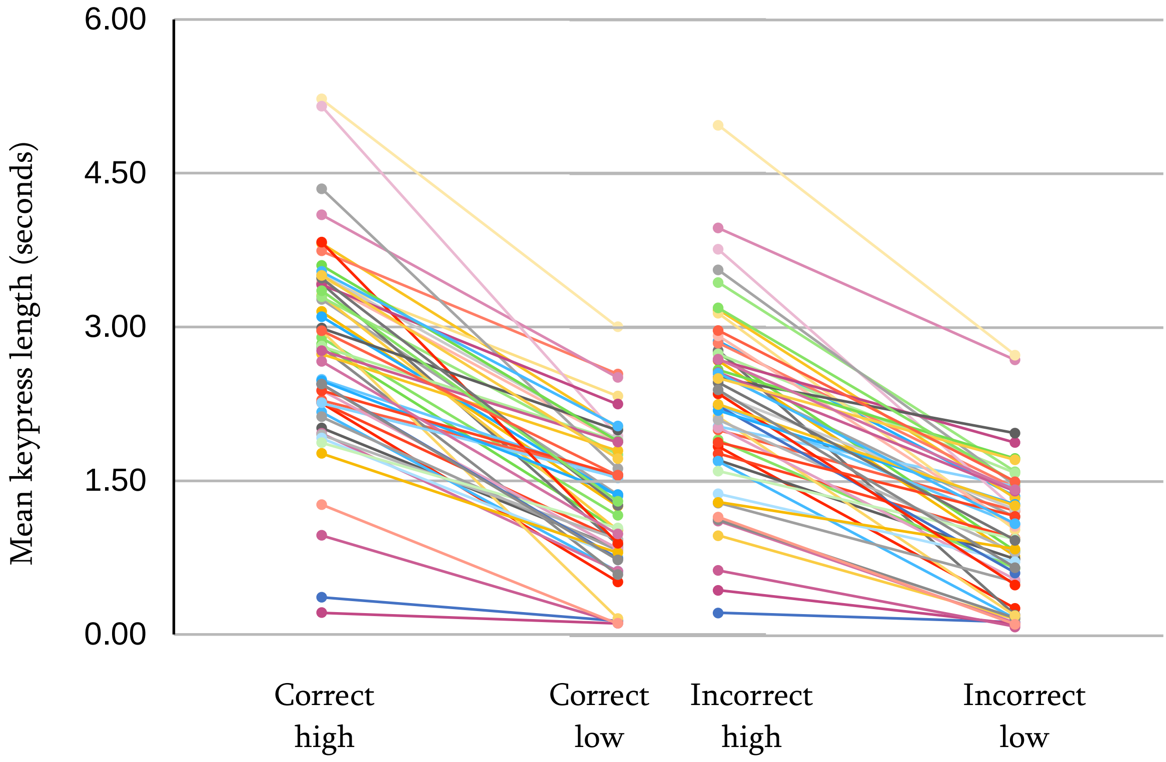

Figure D.2. Validation of high and low confidence conditions for correct and incorrect Test 1 responses in Experiment 3b. Mean keypress lengths (seconds) for Test 1 correct high, correct low, incorrect high and incorrect low conditions.

Appendix E. Supplementary results for Chapter 5

Appendix E.1. Relationship between accuracy and confidence.

One analysis from Experiment 5 assessed the relationship between recognition accuracy and confidence for Test 1, Test 2 repeated and Test 2 baseline conditions to take advantage of the continuous measures of recognition error (using Euclidean distance) and confidence.. Within each condition, Pearson correlation coefficients were calculated for each participant, initially separate for each session, to correlate recognition error (Euclidean distance between a selected face and the target face for each trial) and recognition confidence (keypress length during confidence response for each trial). The average correlation coefficient across EEG session one and two was calculated, before the mean coefficients were calculated across all 40 participants. The mean correlation coefficient values for Test 1 ($t(39) = 2.39$, $p = .01$, $g = 0.38$), Test 2 repeated ($t(39) = 10.44$, $p <.001$, $g = 1.65$) and Test 2 baseline conditions ($t(39) = 9.69$, $p <.001$, $g = 1.53$) were all significantly different to a point value of zero. Furthermore, Bayes factors provided more evidence in favour of the alternative vs. null hypothesis for Test 1 ($\text{BF}_{-0(0.78,0.707)} = 3.17$), Test 2 repeated ($\text{BF}_{-0(0.82,0.707)} = 2.35+10$) and Test 2 baseline conditions ($\text{BF}_{-0(0.46,0.707)} = 2.91+9$) The consistent negative relationship for each condition (see Figure E.1) shows that recognition attempts that were more similar to target face image (i.e. a smaller recognition error value) were related to increased confidence judgements that the selected face was the target (i.e. increased keypress lengths). These results were consistent with those seen in Experiment 3b to indicate that, across the repeated recognition paradigm, recognition responses demonstrated a relationship between recognition accuracy and confidence.

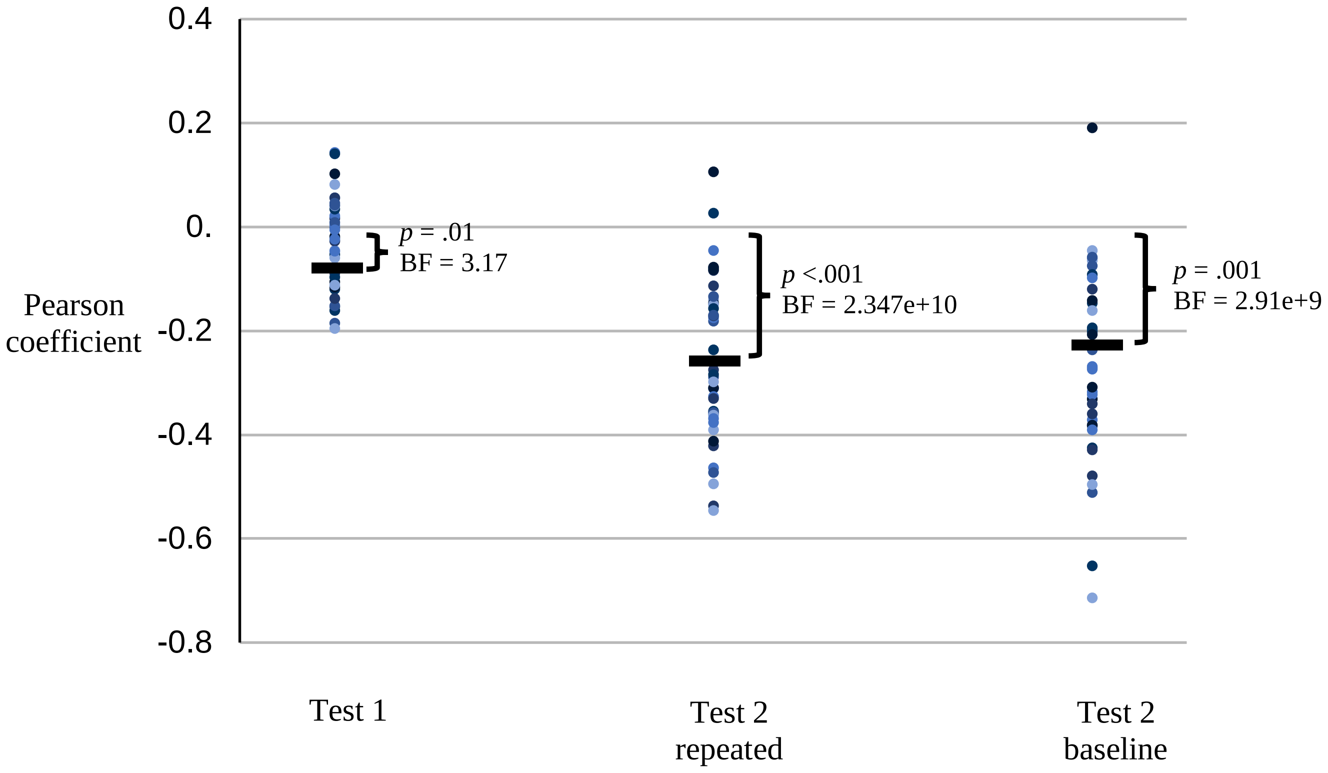

Figure E.1. Correlation between accuracy and confidence across Test 1 and 2. Scatterplot of participant Pearson coefficients analysing the relationship between recognition error and confidence for Test 1, Test 2 repeated and Test 2 baseline conditions. Mean coefficients (black horizontal bars) were significantly different from zero for all three conditions.

Appendix E.2. Supplementary analysis of Test 2 recognition bias using bias difference measure.

The same analyses were conducted on the bias difference measure, which corrected proportion bias scores by the proportion of non-bias response that participants made during Test 2 recognition for “prior error” trials (where a distractor face was selected during Test 1). One-sample t-tests confirmed that mean bias difference measures were significantly greater than zero for Test 1 incorrect high (*M* = 0.16, *SE* = 0.02; *t*(39) = 7.18, *p* <.001) and Test 1 incorrect low conditions (*M* = 0.11, *SE* = 0.02; *t*(39) = 5.07, *p* <.001), indicating that participants were more likely to repeat the same recognition responses across tests rather than switch responses from Test 1 to Test 2. Furthermore, the bias difference measure was significantly greater for the Test 1 incorrect high confidence compared to low confidence condition (*t*(39) = 2.13, *p* = .02, *g* = 0.34), with Bayes Factors confirming more evidence in favour of the alternative vs. null hypothesis ($BF_{10(0.31,0.707)}$ = 2.40), consistent with data from Experiments 3a and 3b.

Appendix E.3. Confirming validity of median split confidence analysis in Experiment 5.

The median split procedure of Test 1 confidence responses, used to create the high and low confidence conditions during Test 1, was next validated to ensure that the null bias effects were not due to a fault with splitting trials into appropriate high and low confidence conditions. As seen in Figure E.2, the data confirm the validity of the median split procedure, as all participants showed higher mean confidence for the high compared to low confidence conditions, for both correct and incorrect trials.

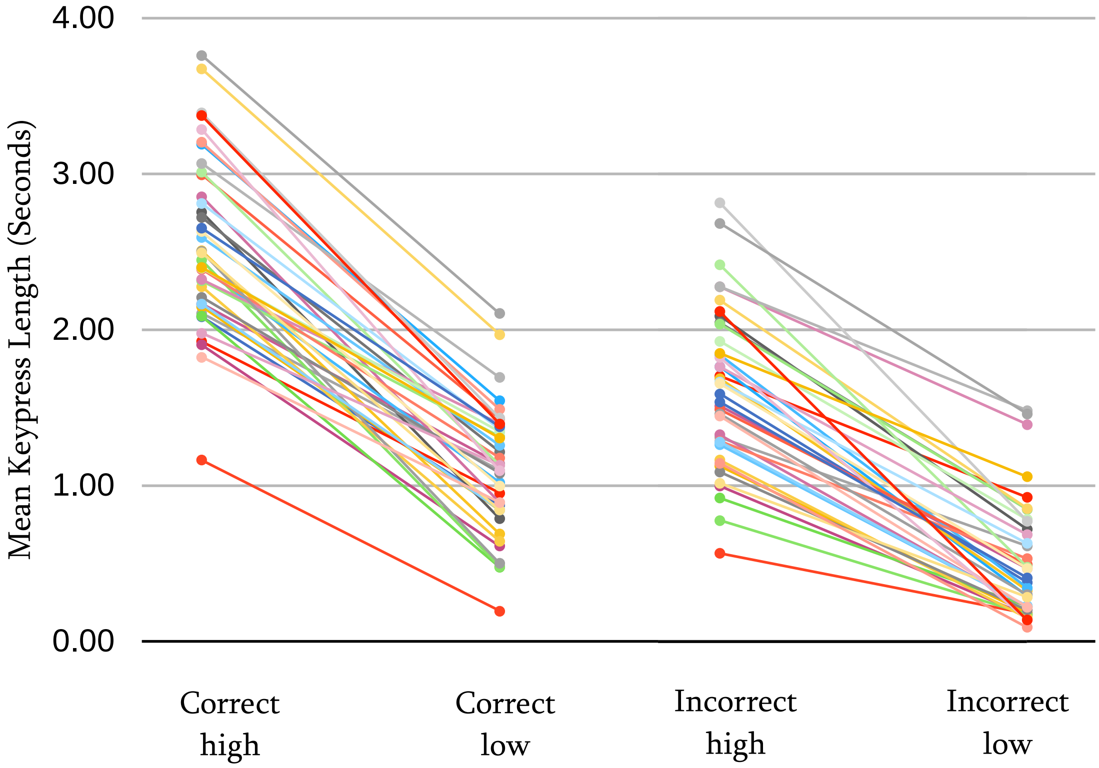

Figure E.2. Validation of high and low confidence conditions for correct and incorrect Test 1 responses in Experiment 5. Mean keypress lengths (seconds) for each participant, for Test 1 correct high, correct low, incorrect high and incorrect low confidence conditions.

Appendix E.4. Statistical analysis of behavioural measures of ERP conditions.

Appendix E.4.1. ERP conditions of accuracy and confidence.

Recognition confidence

Mean confidence was compared between conditions of accuracy and confidence in Test 1 and 2. A 2 (test; Test 1, Test 2) x 2 (accuracy; correct, incorrect) x 2 (confidence; high, low) repeated measures ANOVA was conducted. All main effects and interaction effects within this ANOVA were significant with the exception of the interaction of test and confidence, and the 3-way interaction effect (see Table E.1). Within this ANOVA, it was shown that mean confidence was significantly higher in Test 1 versus Test 2 for the correct conditions (*t*(29) = 6.02, *p* <.001, *g* = 1.10), however there was no significant difference for the incorrect conditions (*t*(29) = 1.98, *p* = .06, *g* = 0.36). Furthermore, mean confidence was shown to be no different between Test 1 and 2 for high confidence conditions (*t*(29) = 1.72, *p* = .10, *g* = 0.32), yet confidence was higher in Test 1 versus 2 for the low confidence condition (*t*(29) = 2.72, *p* = .01, *g* = 0.50). In summary, these results establish that participants were more confident in Test 1 versus Test 2 for the correct condition and the low confidence condition. Finally, analysing the interaction of accuracy and confidence conditions (corrected $\alpha = .016$) showed that mean confidence was significantly higher in Test 1 compared to Test 2 for the correct high (*t*(29) = 4.80, *p* <.001, *g* = 0.87) and correct low conditions (*t*(29) = 4.41, *p* <.001, *g* = 0.80), however no statistical difference was seen for incorrect high (*t*(29) = 2.05, *p* = .05, *g* = 0.37) and incorrect low conditions (*t*(29) = 1.60, *p* = .12, *g* = 0.29). Thus, these findings suggest that participants' confidence for correct responses reduced as a function of repeated testing.

Table E.1. ANOVA results comparing mean confidence for conditions of accuracy and confidence in Test 1 and 2.

ANOVA effects	df	*F*	*p*	η^2p
Test	29	6.87	.014	0.19
Accuracy	29	95.74	<.001	0.77
Confidence	29	390.42	<.001	0.93
Test*Accuracy	29	39.40	<.001	0.58
Test*Confidence	29	0.88	.36	0.03
Accuracy*Confidence	29	25.94	<.001	0.47
Test*Accuracy*Confidence	29	0.13	.73	0.004

For Test 1 conditions of accuracy and confidence, a 2 (accuracy; correct, incorrect) x 2 (confidence; high, low) repeated measures ANOVA showed main effects of accuracy ($F(1,29) = 113.68$, $p <.001$, $\eta^2p = 0.80$) and confidence ($F(1,29) = 464.78$, $p <.001$, $\eta^2p = 0.94$), with mean confidence larger for correct versus incorrect, and high versus low confidence conditions. An interaction effect was also observed ($F(1,29) = 19.51$, $p <.001$, $\eta^2p = 0.40$), with simple main effects analysis showing that mean confidence for high confidence conditions was significantly larger compared to low confidence conditions for both correct ($t(29) = 20.19$, $p <.001$, $g = 3.69$) and incorrect responses ($t(29) = 16.96$, $p <.001$, $g = 3.10$), although the difference as smaller for the latter. Thus, in confirmation of the median split procedure, the average duration of confidence responses in both correct high and incorrect high conditions was significantly longer than those in the correct low and incorrect low conditions respectively, for Test 1 ERP conditions.

A similar pattern of findings was observed for the analysis of Test 2 conditions of accuracy and confidence. A 2 (accuracy; correct, incorrect) x 2 (confidence; high, low) repeated measures ANOVA showed significant main effects of accuracy ($F(1,29) = 58.31$, $p <.001$, $\eta^2p = 0.67$) and confidence ($F(1,29) = 305.94$, $p <.001$, $\eta^2p = 0.91$), with mean confidence larger for correct versus incorrect and high versus low confidence conditions. Finally, a significant interaction was seen ($F(1,29) = 20.21$, $p <.001$, $\eta^2p = 0.41$). Simple main effects analyses in Test 2 showed significantly larger mean confidence for high versus low conditions for both correct ($\mathrm{t}(29) = 16.79$, $\mathrm{p} <.001$, $\mathrm{g} = 3.07$) and incorrect responses ($\mathrm{t}(29) = 15.44$, $\mathrm{p} <.001$, $\mathrm{g} = 2.82$). These findings again confirm a valid median split of confidence responses in Test 2.

Recognition error

Next, retrieval accuracy (operationalised as error, that is the Euclidean distance between the selected distractor face and the target) was analysed to examine the extent to which faces within the incorrect high and low confidence conditions resembled target face images, for both Test 1 and 2. Recognition error for the correct high and low confidence conditions was not analysed as the mean error for these was zero, reflecting the fact that Euclidean values for correct recognition responses is zero for all trials. Analysing recognition error for incorrect conditions of high and low confidence conditions between tests, a 2 (test; one, two) x 2 (confidence; incorrect high, incorrect low) repeated measures ANOVA showed no main effect of test ($F(1,29) = 2.64$, $p = .12$, $\eta^2p = 0.08$), however a significant main effect of confidence was found ($F(1,29) = 18.35$, $p < .001$, $\eta^2p = 0.39$), indicating that mean error was significantly lower for the incorrect high versus incorrect low confidence conditions. However, the lack of any interaction effect ($F(1,29) = 0.08$, $p = 0.79$, $\eta^2p = 0.003$) showed that this difference did not vary as a function of test. Thus, in both tests, participants were more accurate for high compared to low confidence conditions.

The mean recognition error values for the incorrect conditions were both larger than a point value of 3.5 (see Table 5.2 in Section 5.2). In Appendix A, it was established that faces become perceptually dissimilar when the Euclidean distance between images reaches 3.5. For this reason, it can be reasoned that the face images within the two incorrect conditions, both of which had recognition error values significantly greater than 3.5 (incorrect high - $t(29) = 18.69$, $p < .001$, $g = 3.41$); incorrect low - $t(29) = 19.97$, $p < .001$, $g = 3.65$), were perceptually dissimilar to target faces. For Test 2, the mean error for was significantly higher than a point value of 3.5 for incorrect high ($t(29) = 19.71$, $p < .001$, $g = 3.60$) and incorrect low confidence conditions ($t(29) = 25.46$, $p < .001$, $g = 4.65$), showing that on average, the faces selected within the Test 2 incorrect conditions were perceptually distinct from target face images.

Recognition bias

The final behavioural analysis of the accuracy and confidence ERP conditions focused on the Euclidean recognition bias measures across Test 1 and 2. This analysis was conducted to replicate previous results from Experiments 3a and b which had demonstrated that recognition bias is increased for high versus low confidence responses, and specifically to verify if such bias effects were present in the reduced sample and trials used for ERP analyses (compared to the whole-sample, all-trials behavioural analyses presented in previous sections). To this end, mean bias measures were analysed for ERP conditions during both tests 1 and 2. Consistent with the view that confidence and bias are linked, it was expected that mean Euclidean bias should be lower (reflecting smaller Euclidean distances between more similar faces) for the high confidence compared to low confidence conditions in both tests.

In Test 1, a 2 (accuracy; correct, incorrect) x 2 (confidence; high, low) repeated measures ANOVA found a main effect of accuracy ($F(1,29) = 131.77$, $p < .001$, $\eta^2p = 0.82$) and confidence ($F(1,29) = 88.02$, $p < .001$, $\eta^2p = 0.75$). Mean bias distance was significantly lower for correct versus incorrect conditions and high versus low confidence conditions. Furthermore, a significant interaction was seen ($F(1,29) = 4.56$, $p = .04$, $\eta^2p = 0.14$), however simple main effects analysis showed that mean bias distance was significantly lower for correct high versus correct low confidence conditions ($t(29) = 7.74$, $p < .001$, $g = 1.41$) and incorrect high versus incorrect low confidence conditions ($t(29) = 4.30$, $p < .001$, $g = 0.78$). These results confirm that participants included in the ERP analysis were more likely to select a similar face in Test 2 to that selected in Test 1 when Test 1 responses were made with high confidence.

Similarly, for Test 2 conditions, a 2 (accuracy; correct, incorrect) x 2 (confidence; high, low) repeated measures ANOVA showed significant main effects of accuracy ($F(1,29) = 213.13$, $p < .001$, $\eta^2p = 0.88$) and confidence ($F(1,29) = 71.04$, $p < .001$, $\eta^2p = 0.71$). Similar to Test 1, these results showed that mean bias distance was significantly lower for the correct versus incorrect conditions and high versus low confidence conditions. Furthermore, a significant interaction was found ($F(1,29) = 10.10$, $p = .004$, $\eta^2p = 0.26$), however mean bias was significantly lower for correct high versus low confidence conditions ($t(29) = 6.58$, $p < .001$, $g = 1.20$), and incorrect high versus low confidence conditions ($t(29) = 5.31$, $p < .001$, $g = 0.97$). These results show that, during Test 2, participants were more likely to repeat similar recognition decisions to those made in Test 1 that were made with high recognition confidence, both when their decisions were accurate and inaccurate.

Appendix E.4.2. ERP conditions of accuracy and bias

Recognition bias

Firstly, mean bias for the accuracy and bias conditions were analysed to validate the median split procedure used to create the close/far conditions in Test 1 and 2. In Test 1, a 2 (accuracy; correct, incorrect) x 2 (future bias; close, far) repeated measures ANOVA showed significant main effects of accuracy ($F(1,29) = 85.97$, $p < .001$, $\eta^2p = 0.04$) and bias ($F(1,29) = 6411.72$, $p < .001$, $\eta^2p = 0.91$). These results emphasise that mean bias was significantly lower for correct versus incorrect, and close versus far future bias conditions. An interaction effect was also significant ($F(1,29) = 63.19$, $p < .001$, $\eta^2p = 0.01$), with simple main effects analysis showing that mean bias was significantly lower for correct close versus correct far ($t(29) = 59.83$, $p < .001$, $g = 10.92$), and incorrect close versus incorrect far conditions ($t(29) = 48.08$, $p < .001$, $g = 8.78$), however the difference was smaller for incorrect trials. These results validate the median split procedure used to create the close and far bias conditions in Test 1 for both correct and incorrect recognition responses.

For Test 2, similar results were found to those seen in Test 1. A 2 (accuracy; correct, incorrect) x 2 (repetition bias; close, far) repeated measures ANOVA showed significant main effects of accuracy ($F(1,29) = 176.37$, $p < .001$, $\eta^2p = 0.86$) and bias ($F(1,29) = 3794.62$, $p < .001$, $\eta^2p = 0.99$). Here, mean bias was significantly lower for correct versus incorrect, and close versus far repetition bias conditions. A significant interaction also emerged ($F(1,29) = 91.18$, $p < .001$, $\eta^2p = 0.76$), with simple main effects analysis showing mean bias to be significantly lower for correct high versus correct low ($t(29) = 55.08$, $p < .001$, $g = 10.08$) and incorrect high versus incorrect low conditions ($t(29) = 43.83$, $p < .001$, $g = 8.00$), again with a smaller difference between the incorrect conditions. Thus, these results again validate the median split procedure of Euclidean bias values to form ERP conditions of close and far repetition bias in Test 2.

Recognition error

Next, mean Euclidean error was compared for the incorrect close and far bias conditions, for Tests 1 and 2, to assess whether faces within these conditions differed to target face images. A 2 (test; Test 1, Test 2) x 2 (condition; incorrect close; incorrect far) repeated measures ANOVA demonstrated no main effect of test ($F(1,29) = 2.48$, $p = .13$, $\eta^2p = 0.08$), however a main effect of condition was shown ($F(1,29) = 140.11$, $p < .001$, $\eta^2p = 0.83$). Here, mean error was lower for the incorrect close versus far conditions. An interaction between test and condition was observed ($F(1,29) = 7.59$, $p < .001$, $\eta^2p = 0.21$), with post-hoc paired t-tests showing that mean error was not significantly different between Test 1 versus 2 for the incorrect close conditions (*t*(29) = 0.42, *p* = .68, *p* = 0.08), however mean error was lower for the incorrect far condition in Test 1 compared to Test 2 (*t*(29) = 3.08, *p* = .004, *g* = 0 .56). Although mean error was lower for incorrect far condition in Test 1, it is important to reiterate that, on average, the Euclidean distances between target faces and faces in the incorrect far condition was significantly higher than the Euclidean point where faces become perceptually dissimilar in face space (Test 1 incorrect close - *t*(29) = 8.25, *p* <.001, *g* = 1.51; Test 1 incorrect far - *t*(29) = 34.26, *p* <.001, *g* = 6.26; Test 2 incorrect close - *t*(29) = 8.95, *p* <.001, *g* = 1.64; Test 2 incorrect far - *t*(29) = 49.63, *p* <.001, *g* = 9.06). These results indicate that faces within incorrect close and far conditions were perceptually dissimilar to target faces within the respective face sets.

Recognition confidence

Lastly, mean confidence was compared as a function of accuracy and bias conditions between Tests 1 and 2 to again assess whether conditions of bias differed by the average confidence response for trials within these conditions. As previous analyses would suggest, it was expected that close bias conditions should have higher mean confidence than far bias conditions, establishing a link between confidence and recognition bias within the present paradigm. A 2 (test; Test 1, Test 2) x 2 (accuracy; correct, incorrect) x 2 (bias; close, far) repeated measures ANOVA was conducted, with all main effects and interactions shown except for a test x bias interaction, and a test x accuracy x bias interaction (see Table E.2 for inferential statistics). Breaking down the accuracy x test interaction, it was seen that mean confidence was significantly higher in Test 1 versus 2 for the correct condition (*t*(29) = 5.87, *p* <.001, *g* = 1.07) but was higher in Test 2 compared to

Test 1 for the incorrect conditions (*t*(29) = 2.25, *p* = .03, *g* = 0.41). However, for the bias x test interaction, mean confidence between Test 1 and 2 was not different for the close condition (*t*(29) = 1.86, *p* = .07, *g* = 0.34), and between Test 1 and Test 2 for far bias conditions (*t*(29) = 1.72, *p* = .10, *g* = 0.31).

Finally, for the three-way interaction effect, it was confirmed that mean confidence was significantly higher in Test 1 versus Test 2 for correct close (*t*(29) = 6.35, *p* <.001, *g* = 1.16) and correct far conditions (*t*(29) = 3.20, *p* = .003, *g* = 0.58). In contrast, confidence was significantly higher for Test 2 versus Test 1 for the incorrect far conditions (*t*(29) = 2.99, *p* = .006, *g* = 0.55), with no difference in confidence between tests found for the incorrect close conditions (*t*(29) = 1.52, *p* = .14, *g* = 0.28). These findings thus indicate that average confidence for correct recognition attempts reduced as a function of repeated testing, however confidence was elevated during repeated testing for incorrect recognition responses.

Table E.2. ANOVA results comparing mean confidence for conditions of accuracy and bias between tests 1 and 2.

ANOVA effect	df	*F*	*p*	η^2p
Test	29	6.87	.01	0.19
Accuracy	29	95.74	<.001	0.77
Bias	29	390.42	<.001	0.93
Test*Accuracy	29	39.40	<.001	0.58
Test*Bias	29	0.88	.36	0.03
Accuracy*Bias	29	25.94	<.001	0.47
Test*Accuracy*Bias	29	0.13	.73	0.004

Within Test 1, a 2 (accuracy; correct, incorrect) x 2 (future bias; close, far) showed main effects of accuracy ($F(1,29) = 107.25$, $p <.001$, $\eta^2p = 0.79$) and future bias ($F(1,29) = 67.26$, $p <.001$, $\eta^2p = 0.70$). Here, mean confidence was significantly higher for correct versus incorrect and for close versus far conditions. An interaction effect was also observed ($F(1,29) = 24.20$, $p <.001$, $\eta^2p = 0.46$). Post-hoc paired samples t-test's showed mean confidence was higher for close versus far future bias conditions both for correct ($t(29) = 8.57$, $p <.001$, $g = 1.56$) and incorrect conditions ($t(29) = 4.09$, $p <.001$, $g = 0.75$), but the difference was larger for the correct conditions.

Next, for Test 2 conditions of accuracy and repetition bias, a 2 (accuracy; correct, incorrect) x 2 (repetition bias; close, far) repeated measures ANOVA showed main effects of accuracy ($F(1,29) = 54.52$, $p <.001$, $\eta^2p = 0.65$) and confidence ($F(1,29) = 48.94$, $p <.001$, $\eta^2p = 0.63$), with mean confidence larger for correct versus incorrect and close versus far repetition bias conditions. Finally, an accuracy x repetition bias interaction was also observed ($F(1,29) = 24.53$, $p <.001$, $\eta^2p = 0.46$), however post-hoc paired t-test's established the mean confidence was higher for close versus far repetition bias for both correct ($t(29) = 7.03$, $p <.001$, $g = 1.28$) and incorrect conditions ($t(29) = 2.80$, $p = .009$, $g = 0.51$). For both tests, it was thus shown that trials within the correct and incorrect conditions were associated with elevated recognition confidence for trials where participants were biased by future/previous recognition attempts, although this difference was larger between correct than incorrect conditions.

Appendix E.5. Comparing ERPs between Test 1 and 2 for ERP conditions of accuracy and bias.

ERP measures for accuracy and confidence conditions were compared between tests 1 and 2. Cluster corrected paired samples t-tests were performed between tests 1 and 2, with dependent variables of the accuracy-bias interaction (i.e. correct close-correct far bias difference minus incorrect close-far bias difference) as a three-way interaction term. Two-way interactions assessed the difference between correct and incorrect ERP conditions, and close bias and far bias ERP conditions, between tests. Finally, ERPs for the correct close, correct far, incorrect close and incorrect far bias conditions were analysed between tests. Table E.3 shows that no significant clusters for these effects were found, indicating that ERPs for all conditions were similar between Test 1 and 2.

Table E.3. Cluster statistics comparing Test 1 versus Test 2 for accuracy and bias conditions. The largest cluster for each contrast (across positive and negative clusters) is reported.

Cluster contrast	Cluster *t*	*p*
Accuracy x Confidence x Test	-1588.30	.10
Accuracy x Test	669.89	.25
Bias x Test	-1927.30	.06
Correct close-far difference	-1115.20	.15
Incorrect close-far difference	869.43	.20

Appendix F. Supplementary results for Chapter 6

Appendix F.1. Comparing oscillatory power between Test 1 and 2 for accuracy and bias conditions.

The final analyses of the Chapter 6 assessed potential power differences between test for accuracy and bias conditions. In this analysis, for theta, alpha and beta frequency bands, a three-way design was applied with factors of test (Test 1, test 2), accuracy (correct, incorrect) and bias (close, far). For this analysis, cluster corrected paired samples t-tests were performed between Tests 1 and 2 (see Table F.1 for cluster test results), with dependent variables of the accuracy-bias interaction (i.e. correct close-correct far confidence difference minus incorrect close-far confidence difference) as a three-way interaction term. Two-way interactions also compared the difference between correct and incorrect conditions, and close bias and far bias conditions, between tests. Finally, oscillatory power for the correct close, correct far, incorrect close and incorrect far confidence conditions were compared between Test 1 and 2. The only significant cluster was found within the theta band. This effect corresponded to a negative cluster at centro-posterior sites from ~0-0.4s, whereby the difference in theta power between correct and incorrect trials was larger in Test 2 compared to Test 1. This finding indicates that successful target recognition was associated with increased theta power compared to erroneous distractor selection in Test 2 relative to Test 1.

Table F.1. Cluster corrected paired samples t-tests for the 3-way interaction between test, accuracy and bias conditions. The largest cluster (across positive and negative clusters) for each contrast is reported. Contrasts with no clusters formed are denoted as -.

	Theta		Alpha		Beta	
Cluster contrast	Cluster *t*	*p*	Cluster *t*	*p*	Cluster *t*	*p*
Accuracy x Bias x Test	734.75	.14	747.15	.13	179.56	.31
Accuracy x Test	-5497.10	.007	-469.07	.21	-471.43	.11
Bias x Test	-639.89	.21	-82.80	.53	155.85	.40
Correct close-far difference	440.42	.21	-297.06	.28	-53.80	.58
Incorrect close-far difference	-663.20	.19	-378.51	.29	-551.29	.07