Subjective experiences at memory retrieval: The feeling of knowing and beyond

Suzannah Marie Morson

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Institute of Psychological Sciences, Faculty of Medicine and Health

September 2013

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgments

Firstly, I would like to thank all of the older adults and patients who kindly agreed to give up their time to take part in this research. It would not have been possible without them, and it has truly been a pleasure to work with all of the people involved.

I am incredibly grateful to my supervisors, Celine Souchay and Chris Moulin, who have provided me with encouragement and support throughout my PhD. I am especially indebted to Celine Souchay for introducing me to the field of metacognition many years ago. Without her guidance and enthusiasm for my (somewhat overly ambitious) ideas as a wide-eyed undergraduate I would not be where I am today. Throughout my PhD Chris Moulin has also provided invaluable advice and an extra dose of enthusiasm, not only while in Leeds but even from the faraway lands of Dijon. Thanks also to Amanda Waterman for giving me that final push, and Dr Wendy Burn for her invaluable help with participant recruitment. And to Dr Akira O'Connor for an enjoyable month exploring St Andrews and the delights of computational modelling.

I would like to thank my office mates Nathan, Ian and Sarah, for making these last few years so enjoyable. An especially heartfelt thank you to Sarah for her amazing support, for always keeping me smiling, and for providing a never ending supply of chocolate and biscuits to help get us through those tough days. A big thanks to Rebecca, Neil, Anna, Emily, Michelle and everyone else in my year group who have been such lovely and amazing people to know and friends to have. A special mention to Christine, I shall miss our cake-baking exploits, and to Dominika for her calm words of wisdom. And a final thank you to everyone I have failed to mention but who nonetheless have made my time in Leeds truly memorable and fun.

This thesis is dedicated to my parents, for their love, support and belief in me to achieve any goal I set myself; to my brother, who has always brought me back

down to Earth when needed; and to my gran and aunt, who have given me encouragement throughout (and some delicious home-cooking).

Abstract

When we fail to retrieve an item from memory we often experience the sensation that we do know the missing item, it just cannot be recalled right now. Memory is more than retrieval or non-retrieval, it also has a number of sensations and experiences associated with it. The aim of this thesis was to examine subjective experiences at retrieval failure and their association with manipulations of memory accuracy. This was achieved by the use of an existing metacognitive judgement, the feeling of knowing (FOK), and by the development of two novel metacognitive measures, the judgement of retrieval failure (JORF) and the global feeling of knowing (GFOK). In addition to experimental manipulations of memory, these judgements were also examined within populations who typically exhibit memory deficits; healthy older adults and patients diagnosed with early stage dementia. Chapter 2 focused on semantic and episodic FOK in ageing, identifying an age-related selective deficit in episodic FOK accuracy. Chapter 2 also observed that FOK accuracy increased in young and older adults in line with increases in recall accuracy over repeated learning trials. Chapter 3 explored manipulations of retention and retrieval, observing reliable changes in FOK magnitude as recall accuracy was affected, while effects on FOK accuracy were not necessarily in agreement with recall performance. Chapter 4 considered the underlying assumptions of the FOK experience, and proposed a new model of FOK based within the signal detection theory framework. Chapter 5 established two novel measures of retrieval failure, JORF and GFOK. These measures were found to be sensitive to manipulations of memory, and also appear to be preserved in patients with early stage dementia. This thesis provides an important extension to the existing literature on the FOK as well as identifying novel directions for metacognitive theory.

Abbreviations

CAM – Cognitive Awareness Model

DICE – Dissociable Interactions and Conscious Experiences

F - Familiar

FOK – Feeling of knowing

GFOK – Global feeling of knowing

JOC – Judgement of confidence

JOL – Judgement of learning

JOR – Judgement of retention

JORF – Judgement of retrieval failure

K – Know

MCH – Memory constraint hypothesis

MIA – Metamemory in Adulthood

MARS – Memory Awareness Rating Scale

MMSE – Mini Mental State Examination

PFC – Prefrontal cortex

R – Remember

RBMT – Rivermead Behavioural Memory Test

RK – Remember know

SDT – Signal detection theory

TOT – Tip of the tongue

Contents

Figures

Tables

Chapter 1. General Introduction

1.1 Overview

“The universal conscious fact is not “feelings and thoughts exist”, but rather “I think” and “I feel.”” James (1950).

Discussion of the nature of consciousness has historically been the domain of philosophers, dating from Plato and Aristotle through to modern day thinkers. Within the realms of psychology, however, the examination of subjective experiences and awareness were largely avoided. Early work by Wundt on introspection, whereby participants were trained to interrogate their conscious experience and report the constituent sensations and feelings, was later abandoned due to a lack of reliability of such reports and concerns about the self-referent paradox, how the same organ can observe and also observe these observations (Comte, cited in Nelson, 1996, p 103). This led the field of psychology to abandon introspection altogether in favour of the behaviourist approach, where only direct observable behaviour was measured. The rich variety of human experience was largely ignored in order to allow for the clear prediction and control of behaviour.

Gradually, evidence began to accumulate that perhaps the inclusion of subjective reports may provide a more complete view of the cognitive processes engaged during various psychological tasks. For example, Eagle (1967) observed that the accuracy of recall for a word list was related to the strategy participants reported using rather than the strategy they were instructed to use by the experimenter. The effectiveness of mnemonic strategies was therefore more accurately assessed based on subjective reports of strategy use rather than objective instruction of strategy use. Studies such as these marked a turning point in the psychologist’s approach to considerations of conscious experience, allowing the use of participants’ introspections in conjunction with objective measures of behaviour.

Advances in psychological theories and the advent of neuropsychological and neuroimaging techniques have led to consciousness becoming a key topic within the psychological literature. Awareness of our environment and subjective experiences are as important to understanding human behaviour as the behaviour itself.

The aim of this thesis was to explore how factors which affect memory also affect awareness of memory. In particular, this thesis adopted a metacognitive approach to sensations of retrieval failure. This aim was achieved through experimental manipulation of memory at encoding, retention and retrieval, and examining subsequent effects on the sensitivity and accuracy of an established measure, the feeling of knowing, and through the development of a novel measure, the judgement of retrieval failure. In addition, these measures were explored within individuals with memory impairment due to healthy ageing and due to dementia. This work therefore provides new insights into whether deficits in memory accuracy are associated with deficits in metacognitive awareness. Finally, this thesis proposes a novel model to conceptualise the feeling of knowing, which has important implications for the theoretical literature on the basis of metacognitive judgements.

This chapter first presents how the area of research into experiences, sensations and knowledge of cognition has developed, from exploration of the deficits of awareness in patient groups to the experimental approach of metacognition. The metacognitive framework is then considered, with particular focus on the feeling of knowing paradigm. The relationship between memory and the feeling of knowing is then detailed. Findings from experimental manipulations within healthy individuals, pharmacological effects, and consideration of the pattern of deficits observed within neuropsychological populations are reviewed, in addition to findings from non-laboratory based studies. The chapter ends with an overview of the aims of each chapter within the thesis.

1.1.1 Anosognosia

Perhaps the best demonstration of the importance of awareness is within populations where disorders of awareness occur. Babinski (1914) coined the term *anosognosia* to describe a lack of awareness in hemiplegic patients who seemed to ignore their paralysis. The term has since been applied to any disorder in which a lack of awareness is shown regarding symptoms, loss of function and cognitive impairment (McGlynn & Schacter, 1989). Rehabilitation for patients with anosognosia is extremely difficult. Take, for example, an individual who is unaware that they have an impairment in memory. This patient believes that their ability to recall information is the same as it was prior to illness. When presented with a memory task therefore, the patient will adopt the same encoding strategies that they have always relied on, and will expect their performance to be at the same level as usual. The patient will subsequently be disappointed at test when they fail to reach this perceived performance level. In contrast, consider the patient who is aware of their memory problems. During learning, this patient will know that the strategies they have previously relied on are insufficient to allow them to achieve the level of recall they want, and so will change their strategy to encourage more substantial encoding. This change in strategy adoption will subsequently allow the patient to achieve the level of memory performance they wish to at test. By being aware of their impairment, the patient is able to alter their behaviour and their goals in order to allow maximum performance.

A number of studies have identified clear clinical implications of anosognosia. Patients with a lack of awareness as to their impairments have been shown to have low motivation for therapy (Fleming, Strong, & Ashton, 1998), set unrealistic goals (Fleming, Strong, & Ashton, 1996), develop fewer compensatory strategies (Ownsworth, McFarland, & Mc Young, 2000), and to resist support or treatment recommendations. Receptiveness to feedback and a willingness to alter self perceptions have also been identified as essential factors to successful rehabilitation outcomes, factors which require a level of awareness of the difficulties the patient is experiencing (Ezrachi, Ben-Yishay, Kay, Diller, & Rattok, 1991; Ownsworth & Clare, 2006). Thus, the presence of anosognosia can severely impact the outcomes of rehabilitation and limit its success (Clare & Woods, 2004; Ownsworth & Clare, 2006)

Two dominant frameworks have been proposed to describe anosognosia, which are briefly considered here. Within the Dissociable Interactions and Conscious Experience model (DICE; McGlynn & Schacter, 1989), inputs from perceptual, memory and knowledge modules are received by the conscious awareness system, and the processing of these inputs results in the experiences of knowing, remembering and perceiving information. Anosognosia occurs due to disconnection of these modules from the conscious awareness system. This model predicts that different neurological conditions could lead to different modules being disrupted, resulting in domain specific anosognosia. For example, if a memory impaired individual was unable to update their knowledge module, they would show awareness of their memory ability immediately after completing a task but show no awareness of ability within the general context of memory tasks. They would essentially forget that they forget, and so would exhibit mnemonic anosognosia. The Cognitive Awareness Model (CAM; Morris & Hannesdottir, 2004) is a more complex framework to allow consideration of a greater variety in the types of anosognosia that may occur. The model includes comparator mechanisms which monitor existing information and can compare this to the personal data base of beliefs and knowledge about abilities. A metacognitive awareness system receives inputs from both the comparator mechanisms and direct from the personal data base, and can use this to guide behaviour. For example, an individual may believe that they are good at learning faces but not so good at learning names. When presented with a task involving the learning of names and faces, the individual may therefore adopt a different strategy for encoding the names than to that adopted to encode the faces. Anosognosia can therefore occur due to deficits in the metacognitive awareness system, disrupting regulation of behaviour, or at lower system levels such as within the data base, thus providing inaccurate information according to which behaviour is altered.

1.1.2 Metacognition and metamemory

Although the DICE and CAM models provide insight into deficits of awareness, they did not consider these deficits in relation to normal patterns of self awareness. Investigations into the conscious experience of cognitive function within healthy populations led to the arrival of the field of metacognition, which developed independently of the anosognosia literature. Metacognition can be loosely defined as “knowing about knowing”, with metamemory being a specific subclass of metacognition about memory processes. The term metamemory was introduced by Flavell (1971) to encompass the beliefs, attitudes, sensations and knowledge of memory, although the idea of incorporating a person’s own knowledge of their memory function within models of memory had been present in the literature previously (Tulving & Madigan, 1970). A number of researchers subsequently attempted to identify the precise types of knowledge which constitute a person’s metamemory (Brown, 1978; Flavell, 1979; Miyake & Norman 1979; Pressley, Borkowski & Schneider, 1989). For example, Flavell and Wellman (1977) identified three categories of knowledge which could be encompassed within metamemory; person variables, task variables and strategy variables. Person variables include knowledge of how one learns and how one’s memory functions in a general sense. Task variables include knowledge of how task manipulations can affect memory performance, such as the effect of a delay on recall accuracy. Strategy variables include knowledge of the various strategies available for use and how and when to implement them, for example using semantic grouping to improve encoding of mixed word lists.

The focus on taxonomies of knowledge temporarily limited progress within the metamemory field. The types of knowledge that could be construed as being related to successful (or indeed unsuccessful) memory function can never be exhaustive, and so research could become overly concerned with demonstrative studies of the types of knowledge present or absent within certain populations (Cavanaugh & Perlmutter, 1982). A more integrative approach which allowed for theoretical development within the field was provided by the metamemory framework of Nelson and Narens (1990, 1994). The framework outlines the types of memory experiences and sensations which may occur at different stages within the memory process, allowing for consideration of how different tasks may ultimately affect experience and the types of knowledge on which these experiences are based. This framework remains the predominant approach used today, and is the framework adopted by the present thesis.

1.2 The metamemory framework

The core assumption of the Nelson and Narens metamemory framework is that the metacognitive system contains two interrelated levels; the object-level and the meta-level. Memory processes operate at the object-level, while the meta-level contains beliefs and knowledge of how memory functions, as well as an imperfect model of what is occurring at the object-level. Information is exchanged between the two levels via the processes of monitoring and control (see Figure 1.1). Monitoring processes assess the state of the object-level and feed this information into the meta-level, where it can be used to inform the model of the object-level and to update memory knowledge and beliefs. The information in the meta-level can then be used to inform the object-level via control processes, allowing self-regulation of memory behaviour. For example, suppose you wish to make a cake. You read the recipe and see that you are missing six of the ingredients needed, so must go to the shop to buy them. After closing the recipe book you try to remember the six ingredients (i.e. monitoring) but can only recall five of them. You know you cannot make the cake without the correct ingredients, therefore you decide to write a list of what you need and take this with you to ensure you are able to buy everything necessary (i.e. control). Maximal memory performance therefore relies on efficient metamemory function and the feedback loop created by adequate monitoring and control processes (Dunlosky & Connor, 1997).

The direction of the relationship between monitoring and control has primarily been conceptualised as the *monitoring-affects-control hypothesis* (e.g. Koriat & Goldsmith, 1996; Nelson, 1996), whereby subjective experiences affect behaviour.

Dunlosky & Hertzog (1998) proposed a more detailed discrepancy reduction model to explain self-regulation of study time, however the basic principles can be applied to any monitoring-control relationship. During a self-regulated memory task, a desired goal for learning is set, known as the *norm of study* (Le Ny, Denhiere, & Le Taillanter, 1972). Task appraisal and the selection of an initial strategy is based on general metacognitive knowledge at this stage. As learning continues, monitoring of the items of study occurs, and the current memory state of each item is compared to the norm of study. If the item has reached the level required by the norm of study, then encoding of that item is ended. If the item has not yet reached the norm of study, more time is allocated to learning and it is then monitored again. This loop continues until the norm is reached, at which point encoding ends. Deficits in memory performance could relate to deficits at any stage within this loop. For example, an inappropriately low norm of study could be set, leading to insufficient encoding to allow full performance on the task. Alternatively, there could be a deficit in monitoring. Ineffective monitoring leads to poor information being passed to the meta-level, resulting in an imperfect representation of the object-level. This imperfect representation subsequently leads to the incorrect control mechanism being employed, potentially leading to a reduction in memory performance.

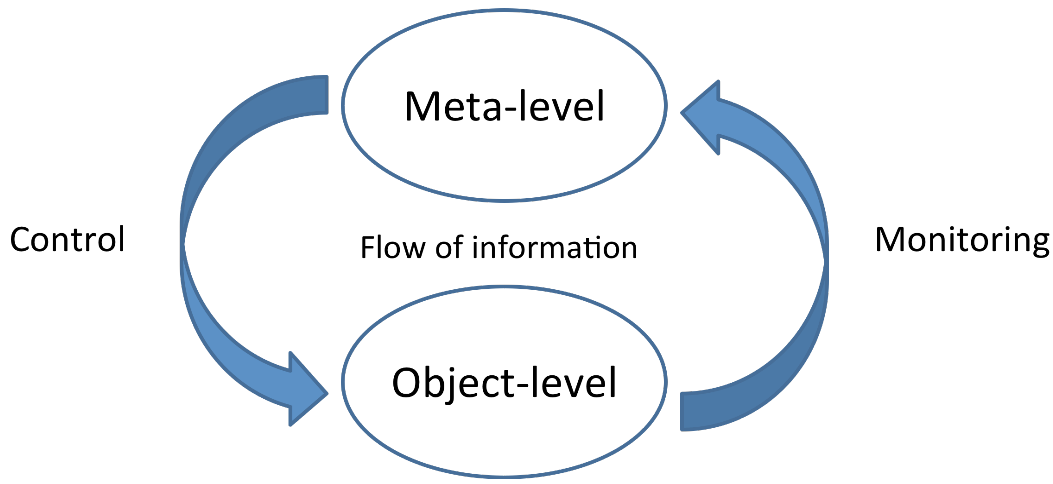


Figure 1.1: The Nelson and Narens (1990; 1994) metamemory framework

1.2.1 Measurement of metamemory

One approach which has been used to evaluate metamemory ability is through questionnaires. These questionnaires typically require self-appraisal of memory functioning on a variety of everyday tasks. Some focus on memory complaints (e.g. the Inventory of Memory Complaints; Herrman & Neisser, 1978), some on memory self efficacy (e.g. the Memory Self Efficacy Questionnaire; Berry, West, & Dennehey, 1989), while others assess metamemory more generally. Perhaps the most widely used of these measures is the Metamemory in Adulthood questionnaire (MIA; Dixon & Hultsch, 1983). The MIA consists of eight scales designed to measure different aspects of metamemory, including use of strategies, awareness of change in memory, locus of control in memory abilities, and memory and state anxiety. These eight dimensions map onto three memory constructs: metacognitive knowledge, memory self efficacy, and memory affect. Similarly, the Memory Functioning Questionnaire (MFQ; Gilewski, Zelinski, & Schaie, 1990) allows evaluation of four memory constructs (general frequency of forgetting, seriousness of forgetting, retrospective functioning, mnemonics usage) via completion of seven scales.

A number of researchers have used questionnaires to examine memory beliefs independently of memory behaviour. These studies have demonstrated that people do not perceive their memory to be a single, unitary construct, but rather hold beliefs about specific skills and situations as well as having a more general sense of their own memory ability (e.g. Broadbent, Cooper, FitzGerald, & Parkes, 1982; Herrman & Neisser, 1978). In addition, memory beliefs have been found to be associated with factors other than memory performance. For example, Broadbent et al. (1982) found correlations between memory beliefs and susceptibility to cognitive failures under stress, while Herrmann (1982) suggests beliefs about memory may also be related to motivational states. Although the use of questionnaires in isolation has provided some insights into general memory beliefs, this only provides part of the story. A point of reference is also required to establish if these beliefs are accurate predictors of actual performance. For example, it is possible for a patient to report that their memory is functioning within the normal range, despite clear deficits on objective memory tests. Within patient groups, many researchers have adopted the approach of using caregiver ratings as a reference point. Parallel versions of the questionnaire are given to the patient and carer, with the patient responding in relation to their self while the carer responds in relation to the patient. The discrepancy between the patient's perception of their difficulties and the carer's perception can then be calculated (Vasterling, Seltzer, Foss, & Vanderbrook, 1995). However, with this type of paradigm it is difficult to differentiate patient overestimation from carer underestimation (Green, Goldstein, Sirockman, & Green, 1993; Trosset & Kaszniak, 1996).

Performance on an objective test of memory performance can also be used as a comparator (Dalla Barba, Parlato, Iavarone, & Boller, 1995; Feehan, Knight, & Partridge, 1991), with memory beliefs assessed either before or following the objective test. This approach raises its own problems, as the memory test used to ascertain performance would need to be matched to the memory beliefs assessed in the questionnaire. As already noted, people do not believe their memory ability is a unitary construct, thus to be informative the subjective ratings must reflect the objective task. With this is mind, Clare, Wilson, Roth, & Hodges (2002) developed the Memory Awareness Rating Scale (MARS) which could be used in conjunction with the Rivermead Behavioural Memory Test (RBMT; Wilson, Cockburn, & Baddeley, 1985), a commonly used neuropsychological test battery. The RBMT is designed to be analogous to a number of everyday memory tasks, such as remembering a name or delivering a message. This provides the MARS with the advantage of being able to obtain measures of beliefs for these tasks in an abstract sense, as most memory belief questionnaires do, as well as obtain measures of beliefs for the specific task the patient has just completed. Caregiver ratings are also gathered to provide a more complete view of the patient's difficulties.

The questionnaire approach to measuring metamemory stems primarily from the anosognosia perspective, looking at subjective reports of how well memory is believed to be functioning generally or in relation to everyday tasks. Within clinical populations these assessments of memory ability are of primary interest, as these inform the clinician as to how the patient is able to function in daily life. The metamemory framework provides an alternative, more experimental, perspective from which to measure metamemory ability. Rather than assess general beliefs about how memory functions, people can be asked to make subjective judgements about their performance on a specific task and this can be then be compared to their actual performance on the task. These judgements can either be global or item-by-item. Global judgements allow a prediction or evaluation of the whole task. For example, Moulin, Perfect, & Jones (2000) asked participants to judge how many items out of 10 they would be able to remember at test. Comparison with the actual number of items recalled allows assessment of whether people over or under estimate their performance. Item-by-item judgements require a judgement or prediction for every item in the task. Subsequent behaviour for each item is then compared to the judgement, and the association between the two examined. Nelson & Narens (1990, 1994) proposed that judgements could be made at each stage of the learning process, at acquisition, retention and retrieval, and that these judgements could either reflect monitoring or control processes (see Figure 1.2). For example, during the acquisition stage, we can monitor how well we are learning an item and predict the likelihood of being able to retrieve it at test, termed a Judgement of Learning (JOL; e.g. Nelson & Dunlosky, 1991). Similarly, predictions about future recognition can be made following unsuccessful recall attempts, called the Feeling of Knowing (FOK; Hart, 1965). The JOL and FOK are both examples of *prospective* memory judgements, a prediction of what future memory performance will be. Alternatively, judgements can be *retrospective*, an evaluation of how well the person feels they have completed the task, for example a confidence judgement about performance. This thesis focuses on the FOK, and so this will be presented in more detail below.

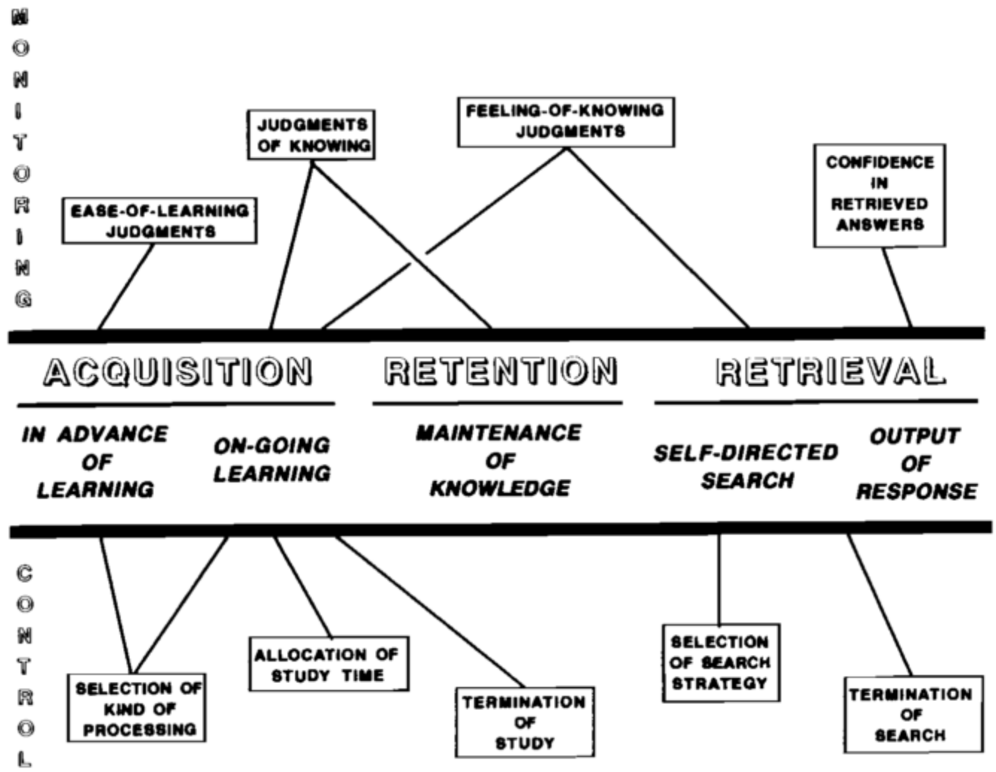


Figure 1.2: Nelson and Narens (1990; 1994) framework for metacognitive judgements

1.3 The feeling of knowing

First, it is important to note that the FOK literature is driven by memory theory. As the FOK is a sensation associated with a failure to retrieve a memory, it is dependent on the conceptualisation of memory used; advancement of the FOK literature therefore relies on memory theory. That is not to say that the distinctions and processes outlined in memory have metacognitive equivalents. Indeed, as will be discussed later, memory and FOK are not always found to be closely related. Nonetheless the memory literature provides the framework in which the FOK literature can develop. The central memory theory explored within the thesis will be briefly mentioned here, with further development throughout the relevant chapters as indicated.

A primary distinction within memory theory is that of semantic memory and episodic memory. Semantic memory contains knowledge and facts, while episodic memory contains singular events associated with a particular time and place (Squire, 2004; Tulving, 1972, 1985a). Memory failures can therefore occur for both semantic and episodic information. The question then arises as to whether an FOK for semantic memory is the same as an FOK for episodic memory, a point which we will return to later in the current chapter and in Chapter 2. Semantic and episodic memory are further distinguished by the sensations that are associated with *successful* retrieval. Tulving (1985b) proposed that episodic memory was associated with recollective experience, a mental reliving of the experience due to retrieval of thoughts, feelings and sensations that were encoded in conjunction with the memory itself. In contrast, semantic memory does not have this rich sensation associated with successful retrieval; the rememberer simply 'knows' they have retrieved the correct memory. This distinction has led researchers to consider whether these subjective experiences associated with successful retrieval may be associated with the FOK sensation during unsuccessful retrieval (Hicks & Marsh, 2002; Souchay, Moulin, Clarys, Taconnat, & Isingrini, 2007), a link which will be explored further within Chapters 2 and 3. Thus as research into recollection has advanced, so too has research into its potential role within the FOK.

The influence of further memory theories is considered within individual chapters, as these have invariably determined the design of the task used. The remainder of the current chapter will focus on the FOK paradigm, the basis of FOK judgements, and insights from its use in different aspects of experimental research.

1.3.1 Method

The FOK is a prospective memory judgement made following a failed retrieval attempt. The participant is asked to predict the likelihood of retrieving the item at a later test despite being currently unable to retrieve it. Hart (1965) developed the recall-judgement-recognition (RJR) paradigm to measure FOK, whereby a recall attempt is made, a judgement about the likelihood of future recognition, followed by an actual recognition test. Hart originally used a semantic memory task by giving participants general knowledge questions, for example “Which planet is the largest in our solar system?”. During a semantic FOK task, if participants are unable to recall the correct answer, they are then asked to indicate whether they think they would be able to choose the correct answer from a number of alternatives. A ‘Yes’ response thus reflects a positive FOK, while a ‘No’ response reflects a negative FOK. The same questions are then presented as a multiple choice recognition task, with participants asked to choose an answer for each question. The accuracy of the FOK is then measured using the Goodman-Kruskal gamma correlation (Nelson, 1984), which represents the correspondence between the FOK judgement and later recognition performance. The gamma correlation gives a score between -1 and +1, with large positive numbers indicating high FOK accuracy, scores of zero indicating no predictive accuracy, and negative values indicating an inverse relationship between FOK and recognition (see Chapter 4 for a more in depth discussion of alternative measures used to assess FOK accuracy).

This paradigm has also been applied to episodic tasks (e.g. Hart, 1967), where a learning stage is added prior to the recall, judgement and recognition stages. The participant is typically asked to encode a series of paired associates. At recall, the first word of the pair (the cue) is given, and participants are asked to recall the target word which was presented with it at encoding. If recall is unsuccessful, an FOK judgement is made in the same way as the semantic task, followed by a multiple choice recognition task. Again, the correspondence between judgement and recognition allows measurement of the accuracy of the FOK prediction (see Chapter 2 for a more in depth consideration of semantic and episodic FOKs).

In addition to considering the basic RJR paradigm within various memory tasks, researchers have also varied the RJR itself. For example, the majority of researchers have followed the original Hart methodology and used recognition as the criterion test, as is the case within the current thesis. And yet the precise nature of the recognition task can be altered. Chapter 2 and Chapter 3 use four alternative forced choice recognition, while Chapter 5 uses an old/new recognition task. Other criterion tests can also be used, including cued recall, where the first letter of the target has been provided (Gruneberg, Monks, & Sykes, 1977; Gruneberg & Monks, 1974), reminiscence (Gruneberg & Sykes, 1978; Read & Bruce, 1982), perceptual identification (Nelson, Gerler, & Narens, 1984), lexical decision (Connor, Balota, & Neely, 1992), and stem completion (Lupker, Harbluk, & Patrick, 1991). Likewise, the materials used at encoding for the episodic task have included general information questions (Nelson & Narens, 1980), the names of entertainers (Read & Bruce, 1982), word definitions (Allen-Burge & Storandt, 2000), word translations (Peynircioğlu & Tekcan, 2000), trigrams (Blake, 1973), sentences (Shimamura & Squire, 1986), face-name associations (Hosey, Peynircioğlu, & Rabinovitz, 2009), and various types of paired associates (Leonesio & Nelson, 1990; Nelson, 1988; Souchay et al., 2007). The present thesis used a variety of stimuli, including general knowledge questions (Experiment 2.1), French-English translations (Experiment 2.2), unrelated paired associates (Experiments 3.1 and 3.2), arrays of words, numbers and symbols (Experiments 3.3 and 3.4), and categories (Experiments 5.1, 5.2. and 5.3). In addition, the FOK judgement itself can be varied from a basic Yes/No distinction to rating scales of various lengths, providing participants with the opportunity to give finer grained judgements. Both approaches were used within the current thesis (see Chapter 4 for consideration of the implications of each type of FOK measure).

Although not directly considered in this thesis, it is worth noting that while the FOK occurs at recall, the criterion test is recognition. Multiple areas of research suggest that recall and recognition rely on different systems or processes. For example, within the ageing literature recall has been shown to be impaired in older adults while recognition memory is relatively preserved (e.g. Botwinick & Storandt, 1980; Craik & McDowd, 1987). A number of theories have been proposed to explain how recall processes differ from recognition processes, from trace strength (MacDougall, 1904) to generate-recognise (Anderson & Bower, 1972) to encoding specificity (Tulving, 1983). The current predominant approach is that recall relies on declarative processes, or explicit and conscious memory retrieval, while recognition memory relies on both declarative and non-declarative processes (Haist, Shimamura, & Squire, 1992). Thus in recognition, non-declarative factors may also influence the success of retrieval, such as priming (Gardiner, 1988; Jacoby, 1983; Mandler, 1980). One aspect these theories share is the idea that recall is more difficult and more effortful than recognition.

The differences between recall and recognition processes could therefore be considered an important component of the conceptualisation of the FOK. In the current thesis, the FOK is based on the products of a failed recall attempt, but is a prediction of a future recognition test. The magnitude of the FOK judgement may therefore depend on the participant knowing that recall and recognition are reliant on different processes and that items which are not available at recall may be available at recognition. The FOK is in essence an attempt to quantify that difference in availability. If the participant fails to quantify this difference, this will subsequently affect the FOK accuracy. Thus manipulations of memory which affect recall but not recognition (or indeed the effect of ageing on recall and not recognition) can provide novel insights into how aware participants are of these differences and into the basis of the FOK judgement.

1.3.2 Feeling of knowing and tip of the tongue

The FOK paradigm cannot be considered without also mentioning the tip of the tongue (TOT) literature. A TOT experience, like the FOK, occurs following retrieval failure. The key difference between the two concepts is the sensation of immediacy that accompanies the recall failure during a TOT. While an FOK is an estimate of the likelihood of future recognition, a TOT is an intense feeling that the unrecalled information is known and that retrieval is imminent (Brown & McNeill, 1966). Despite the similarity in the definitions of each sensation, the two literatures have developed largely independently of each other (Schwartz, 2002). While the FOK has been the key paradigm for memory researchers and the focus has been on the accuracy of judgements, the TOT has been considered more in a linguistic context, used to investigate word retrieval (e.g. Burke, MacKay, Worthley, & Wade, 1991; Jones, 1989; Meyer & Bock, 1992; although see Schwartz & Frazier, 2005 for a metacognitive approach). Nonetheless, a number of researchers conceptualise the TOT as a strong FOK, assuming that the two measures reflect the same underlying cognitive process (Bahrick, 2008; Gardiner, Craik, & Bleasdale, 1973; Kozlowski, 1977).

One approach to measuring TOTs is to rely on their spontaneous occurrence in everyday memory. These naturalistic examinations of the TOT experience rely on diary completion by participants over long periods of time, typically between one week and one month (Burke et al., 1991; Burke, Worthley, & Wade, 1988; Sunderland, Watts, Baddeley, & Harris, 1986). Alternatively, TOTs have been elicited in an experimental setting, using a similar method to the FOK. Semantic tasks have predominantly been used following the methodology of Brown & McNeill (1966), although episodic tasks in the form of paired associates have also been examined (e.g. Ryan, Petty, & Wenzlaff, 1982). It is these laboratory based TOT experiences that have been considered when trying to disentangle whether or not TOT and FOK reflect the same cognitive process.

The effect of experimental manipulations on TOT and FOK provide one avenue of exploring the potential links between the two experiences. If an experimental manipulation has the same effect on both subjective states, it may indicate that they do reflect the same process. However, if an experimental manipulation leads to distinct effects, it may be more appropriate to consider the FOK and TOT as separate experiences. Metcalfe, Schwartz, & Joaquim (1993) varied the familiarity of the cue word and the memorability of the target word in a series of four experiments. In each case, the effects on FOK and TOT were the same; the memorability of the target items had no effect on either judgement while more familiar cues led to increases in FOK and TOT. In contrast, Widner, Smith, & Graziano (1996) found that by changing participants' expectations as to the difficulty of the task, varying effects were found on TOT and FOK. Participants who expected easy questions reported a higher frequency of TOTs than those who expected difficult questions, whereas an equivalent number of FOKs were reported by both groups, regardless of the expected difficulty level. Schwartz (2008) manipulated working memory load during a semantic memory task by asking participants to remember four digits while attempting recall for half of the general knowledge questions. Compared to those questions attempted without a concurrent working memory task, participants reported fewer TOTs while FOKs increased. This was despite no effects of working memory load on recall performance. A second study using rehearsal of six digits again showed a decrease in TOT numbers while FOK responses were equivalent for items with and without a working memory component. Using a rare word definitions task, Yaniv & Meyer (1987) asked participants to report first if they experienced a TOT, followed by a rating of FOK strength. In some instances FOK ratings were high in the absence of a TOT, questioning the assumption that TOTs are merely strong FOKs. These studies therefore suggest that TOT and FOK relate to two separate subjective experiences which warrant independent consideration.

The presence of a dissociation between FOK and TOT has also been examined using neuropsychological and neuroimaging methods. During a semantic memory task, Maril, Wagner, & Schacter (2001) identified three key regions of heightened activation associated with TOT experiences; anterior cingulate cortex, right dorsolateral prefrontal cortex, and right inferior prefrontal cortex. In an examination of FOKs with an episodic memory task, Maril, Simons, Mitchell, Schwartz, & Schacter (2003) observed increased activation in left prefrontal regions and parietal regions, with no indication of overlap with areas identified during TOTs. The assumption of independence between FOK and TOT would be premature based on these two studies alone, as the different memory tasks and different stimuli used may be contributing factors to the different patterns of activation recorded. To clarify this, Maril, Simons, Weaver, & Schacter (2005) examined both FOK and TOT within the same task. General knowledge question were used as stimuli, and participants were asked to indicate whether they experienced a TOT or an FOK for retrieval failures. TOT experiences were found to be associated with activation in the anterior cingulate, right dorsal prefrontal cortex, and right inferior prefrontal cortex, as previously. However, no unique activation was identified for FOKs.

The precise role of the prefrontal cortex in each judgement is perhaps confused slightly by the findings of Widner & Winkelman (2005). Here, patients with deficits in prefrontal cortex (PFC) function were found to have associated deficits in FOK but no deficits in TOT. Based on the neuroimaging results with healthy participants which indicated a greater level of activation in prefrontal cortex for TOTs and not FOKs, it would be expected that patients with PFC deficits would show impairment in TOT rather than FOK. Regardless of this slight incongruence, if TOTs are essentially strong FOKs and reflect the same underlying cognitive process as suggested, then the same areas of brain activation should be identified for each judgement and patients would show similar deficits whether asked to make a TOT or FOK. Experimental manipulations should also produce similar effects on both judgements. This has clearly been shown not to be the case, thus FOK and TOT would appear to measure qualitatively different experiences.

1.3.3 The basis of feeling of knowing judgements

Nelson et al. (1984) identified 12 possible mechanisms that may underlie FOK judgements, falling broadly into two categories; trace-access and inferential. Trace access mechanisms assume explicit direct access to the unrecalled target, thus the FOK judgement is based on the strength of the memory trace relating to the exact target in question (e.g. Eysenck, 1979; Gardiner et al., 1973; Gruneberg, 1983). Variables affecting memory performance should therefore have direct consequences on FOK judgements. For example, manipulations which increase memory strength should also increase FOK judgements and manipulations which decrease memory strength should decrease FOK judgements. However, Koriat (1993) demonstrated that FOK judgements for trigrams were more closely related to the number of letters retrieved than to whether the letters retrieved were correct or not. Retrieval of two incorrect letters of an unrecalled trigram could lead to higher FOK judgements than retrieval of one correct letter of a trigram (see also Blake, 1973). If the FOK was based on trace access mechanisms, only correct information should affect the judgement given. Further investigations of the effects of memory manipulations on FOK judgements have failed to demonstrate consistent effects (see Section 1.3.4), leading to inferential mechanisms to be favoured over direct-access views.

Inferential mechanisms rely on heuristics to infer the likelihood of retrieval. The strength of the target memory itself is not of primary concern within these theories, but rather the monitoring of aspects related to the memory. As no direct link between memory and FOK occurs, it is possible for experimental manipulations to have independent effects on each process. Although a number of inferential mechanisms have been proposed, the cue familiarity hypothesis and the accessibility account have received the most support and have become the predominant theories in the literature. Due to the widespread acceptance of these theories and their influence on FOK research, each will be described in more detail below.

Cue familiarity

The cue familiarity hypothesis was first proposed by Reder (1987). Here, the primary driving force behind the FOK judgement is the familiarity of the pointer used to access the target memory. For example, in a paired associates task it is the participant's familiarity with the cue word that would contribute to the FOK judgement rather than any indirect access or evaluation of the target itself. The assessment of cue familiarity is proposed to be an extremely quick process, leading to rapid FOK judgements which enable the individual to decide whether a search of memory is worthwhile or not.

The cue familiarity account originated in studies of the effect of priming. By increasing participants' exposure to cues prior to the recall task, the familiarity of the cue can be manipulated, leading to effects on the FOK. For example, Reder (1987) asked participants to rate the frequency of words, some of which were later presented in general knowledge questions. Those questions which contained cue words that had been rated for frequency received higher judgements of future retrievability than those questions where no words had been rated for frequency. Importantly, no effect was found on actual recall accuracy. Schwartz & Metcalfe

(1992) used a similar paradigm, with participants asked to make pleasantness ratings on half of the cue words used in a later paired associates task. Again, no effect of priming was found for recall, however FOK judgements were significantly higher for unrecalled items where the cue word had been primed than those where no priming had occurred. Manipulations of the accessibility of the target were not found to affect the FOK.

Reder & Ritter (1992) further demonstrated the effect of priming with arithmetic problems. In a speeded FOK task, participants were asked to judge their ability to retrieve the answer to a series of math problems (e.g. 34 x 67), some of which were repeated several times. These rapid (less than 850 ms) FOK judgements were found to be higher for those questions which shared operands with previous problems (e.g. 34 + 67) than those which were unique (e.g. 86 - 29), despite no difference in participants' ability to answer the problem. By repeatedly exposing the participant to the operands it created a feeling of familiarity with similar looking problems, leading to higher FOKs. Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino (1997) further showed a relationship between the frequency of prior exposures and the magnitude of FOK ratings. As the number of exposures increased, so too did the magnitude of the FOK judgements.

Accessibility

The accessibility account (Koriat, 1993, 1995) proposes that FOK judgements are based on the accessibility of information relevant to the target. It is not assuming direct-access to the target itself, but that the failed memory search for the target activates relevant partial information from which a judgement can be made. This partial information can range from vague details about the target in question, for example whether it was a short or long word, or a positive or negative word, through to partial retrieval such as giving the first letter of the target or words with a similar semantic meaning (e.g. Blake, 1973; Eysenck, 1979; Koriat, Levy-Sadot, Edry, & de Marcas, 2003; Koriat, 1993; Schacter & Worling, 1985; Thomas, Bulevich, & Dubois, 2011). The quantity and quality of the partial information retrieved determines the magnitude and accuracy of the FOK judgement (Koriat, 1993).

A number of studies support the accessibility account of FOKs. Koriat et al. (2003) demonstrated that semantic attributes were still available for retrieval despite an inability to recall the target. Participants were able to correctly rate unrecalled items on three semantic dimensions (good-bad, strong-weak, active-passive), indicating some level of access to information relevant to the unrecalled target. Blake (1973) and Koriat (1993) found that partial recall of trigrams led to higher FOK judgements than when no part of the trigram could be retrieved. Interestingly, this increase in FOKs occurred whether the partial recall was correct or incorrect, demonstrating the influence of the quantity of partial information available. The more details retrieved that are thought to be correct, the higher the FOK. The importance of the quality of information was demonstrated by Thomas et al. (2011), who observed that FOKs were higher for unrecalled items when the correct valence was retrieved as opposed to unrecalled items where the incorrect valence was reported (see also Schacter & Worling, 1985; Thomas, Bulevich, & Dubois, 2012).

Integrated account

Given the amount of experimental support for both the cue familiarity and the accessibility accounts of FOKs, an integrated account has been proposed by combining the two theories (Koriat & Levy-Sadot, 2001). Cue familiarity and accessibility are proposed to be two separate mechanisms which affect the FOK in a cascaded manner. When initially presented with a recall task, a rapid preliminary FOK is formed based primarily on the familiarity of the cue. The level of FOK reached determines whether a search for the target is initiated or whether the participant reports that they do not know the answer (Reder, 1987). If this initial FOK is high enough to instigate a memory search, the amount of accessible information retrieved during this search is then integrated into the FOK, updating the judgement made. If the initial FOK is low, then no memory search will occur, and so accessibility information will make a minimal contribution to the FOK judgement reported (Koriat & Levy-Sadot, 2001).

The above evidence cited for each individual mechanism can easily be explained with the integrated account. The priming studies used to support the cue familiarity account involve rapid FOK judgements, thus the participant does not have sufficient time to initiate the memory search and allow accessibility information to affect their judgement. Only manipulations of cue familiarity are able to influence the FOK within this paradigm, as manipulations of the target require more time to access and evaluate and subsequently inform the FOK (Schwartz & Metcalfe, 1992). Due to the cascaded nature of the proposed integrated theory, the findings from the accessibility account do not require further interpretation, as sufficient time is allowed for a retrieval attempt to be made and the effects of partial information to be incorporated into the final FOK. It does, however, allow for cue familiarity to also be considered as an aspect of the partial information on which the FOK is based. Importantly, although cue familiarity contributes to the decision to search memory for the target, following memory search it is not necessarily a major contributor to the reported FOK judgement. The two processes may interact, or may operate independently, allowing for different mechanisms to be employed in different circumstances (Miner & Reder, 1994).

1.3.4 The link between memory and FOK

Theories of direct access to target memories have been discounted as the basis of FOK judgements. However, consideration of the effects of memory manipulations on FOK judgements can provide useful insight into the cognitive processes which may or may not be associated with the FOK. Indeed, it is the aim of this thesis to explore manipulations of memory at encoding, retention and retrieval. Furthermore, understanding of these effects in healthy populations and assessment of the metacognitive abilities of various patient groups enables identification of potential deficits in awareness. The current section will briefly consider experimental manipulations of memory in healthy participants and findings from the neuropsychological literature.

FOK accuracy is judged in terms of the relationship between the judgement and subsequent memory performance, with the finding that participants are generally accurate in their predictions (e.g. Gruneberg & Monks, 1974; Kelemen, Frost, & Weaver, 2000; Nelson, Leonesio, Landwehr, & Narens, 1986; Souchay, Isingrini, & Espagnet, 2000). As already discussed, part of the information used when making an FOK is the amount of partial information accessible about the target memory, thus indicating an indirect link between memory and FOK. Indeed, if FOK judgements were not predictive of future memory accuracy, the subjective experience of knowing an answer without direct retrieval would hold no cognitive advantage. Of equal interest is the relationship between memory traits *prior* to the FOK judgement and subsequent FOK predictions. Does an increase in recall performance for a task provide a better basis on which FOK judgements can be made? Do interventions which impair memory also impair FOK?

1.3.4.1 Manipulations in healthy participants

Young adults

With regards to information processing, memory consists of three stages: encoding, retention and retrieval. Experimental manipulations can occur at any of these stages and lead to effects on recall accuracy, without increasing or decreasing participants' ability to retrieve the correct information. Schacter (1983) employed a series of classic memory manipulations to explore their effects on FOK for paired associates at each stage of memory. Experiment 1 examined the effect of delay. Participants were tested on half of the pairs immediately after encoding, as in the standard FOK paradigm. The remaining items were tested one week later. Typical memory effects were found, with recall accuracy (and subsequent recognition) lower following a week delay than when tested immediately. However, no effect was found on FOK accuracy. A second experiment manipulated presentation time, with the finding that although shorter presentation times led to lower recall accuracy, no effect was found for FOK accuracy. A final experiment looked at the match of the cue used at encoding and retrieval. At recall, the cue word given to elicit the target was either identical to that during encoding, rhymed with the cue word, or was semantically related to the cue word. Recall accuracy was significantly higher when the identical cue word was given, and FOK accuracy was also significantly higher in this condition.

From Schacter's (1983) observations it may be considered that manipulations of encoding and retention of memory have no effects on FOK accuracy, while manipulations at retrieval can alter the accuracy of the judgements made. Indeed, in Hart's early work on the FOK he also found that manipulating the repetition of items at encoding had no effect on FOK accuracy despite clear benefits to recall (Hart, 1967). However, a number of later findings suggest this is too simplistic. Sacher, Taconnat, Souchay, & Isingrini (2009) asked participants to encode paired associates either under full attention or divided attention conditions. The items encoded under divided attention showed lower recall accuracy than those encoded under full attention, and FOK accuracy showed a similar pattern. FOK judgements for unrecalled items which had been encoded under full attention were more accurate than those for unrecalled items encoded under divided attention. In a manipulation to improve memory accuracy, Carroll & Nelson (1993) and Nelson, Leonesio, Shimamura, Landwehr, & Narens (1982) both demonstrated that overlearning of items led to higher recall accuracy and higher FOK accuracy.

Older adults

In the same way that memory theory has driven the FOK literature, so investigations into healthy ageing have driven the memory literature. Research has suggested that ageing does not lead to a general decline in memory ability. Rather, some aspects of cognitive performance appear to be preserved, while other aspects are impaired (Craik, Anderson, Kerr, & Li, 1995). These changes in performance across tasks therefore allow for greater understanding of memory function, and also provide support for the taxonomic distinctions used within the memory literature (Craik et al., 1995). Thus changes in cognition during healthy ageing also provide a valuable tool for examining metacognitive function. Older adults provide a unique opportunity to investigate the effects of naturally occurring deficits in memory on metacognitive judgements. If impairments are also observed in metacognition, the question then arises as to whether this impairment is due to a genuine deficit within metacognitive processes or is a consequence of a deficit in memory performance.

One of the central distinctions of memory theory is between semantic and episodic memory (Tulving, 1972), with semantic memory referring to factual knowledge while episodic memory is linked to a specific time and place. In healthy ageing, episodic memory declines while semantic memory remains intact (Anderson & Craik, 2000; Zacks, Hasher and Li, 2000). However, while semantic FOK is preserved in ageing (Butterfield, Nelson, & Peck, 1988; Marquié & Huet, 2000), no consensus has been reached on the effects on episodic FOK, with evidence for both preservation (MacLaverty & Hertzog, 2009) and impairment (Perrotin, Isingrini, Souchay, Clarys, & Taconnat, 2006; Souchay et al., 2000; Thomas et al., 2011). Two competing hypotheses have attempted to account for these contradictory findings. The memory constraint hypothesis (MacLaverty & Hertzog, 2009) proposes that observed deficits in episodic FOK accuracy are a direct consequence of deficits in episodic memory accuracy, and so can be removed by improving memory ability in older adults. In contrast, Souchay et al. (2007) and others propose that older adults show a deficit in episodic FOK which is independent of their impairments in episodic memory. These two theories will be presented and discussed in more detail in Chapter 2.

This brief overview of memory manipulations in healthy populations shows the complex nature of the relationship between memory and FOK. Manipulations at each stage of the memory process can have varying effects, indeed even similar paradigms such as repeated presentation (Hart, 1967) and overlearning (Carroll & Nelson, 1993; Nelson et al., 1982) do not necessarily have similar influences.

1.3.4.2 Pharmacological manipulations

The effects of a number of pharmacological compounds have been considered during studies of the FOK. Although these agents tend to have quite widespread and diverse effects on the brain, they can still provide information on cognitive function. For example, alcohol affects cell function in the cerebellum (Ming, Criswell, Yu, & Breese, 2006), leading to problems with motor function and co-ordination. Acute alcohol administration also impairs semantic memory and episodic memory, primarily when administered during encoding but can also have effects at retrieval (Mintzer, 2007; Söderlund, Parker, Schwartz, & Tulving, 2005). Nelson, McSpadden, Fromme, & Marlatt (1986) used a semantic memory task to investigate the effects of alcohol on FOK. Despite clear deficits in recall of general knowledge information, no effect was observed on FOK accuracy. Alcohol was found to impair memory without impairing monitoring of future memory performance.

Benzodiazepines in particular have been a focus of metacognitive research due to their profound amnesic effects (Bacon, Schwartz, Paire-Ficout, & Izaute, 2007; Schwartz & Bacon, 2008). Impairments of episodic memory are universal to all drugs under the benzodiazepine classification, although their effects on semantic memory may vary from drug to drug (Buffet-Jerrott & Stewart, 2002). Lorazepam and midazolam have been examined within the FOK paradigm (Bacon et al., 1998; Izaute & Bacon, 2006; Merritt, Hirshman, Hsu, & Berrigan, 2005). Merritt et al. (2005) focused on the effects of midazolam in a standard paired associates task. Deficits were observed in episodic memory accuracy, however no effect was found for FOK accuracy. Despite impairments in recall participants were still able to predict their future recognition performance. Bacon et al. (1998) considered the effect of lorazepam for episodic and semantic memory and FOK judgements. Clear impairments of memory were found for both sentence learning and retrieval of general information. With regards to the FOK, some evidence of a deficit was observed in episodic FOK accuracy. Participants who received a dose of lorazepam were found to be at chance accuracy with their FOK predictions, while participants administered with a placebo were reliably above chance. However, no significance difference was detected between the two groups, thus it cannot be concluded that episodic FOK was impaired. No effect of drug was observed on the semantic FOK accuracy.

Izaute & Bacon (2006) further explored the effects of lorazepam on episodic memory and FOK with the additional assessment of the effects of partial information. Four letter nonsense strings were presented at encoding, allowing participants to recall either the full target item or individual letters from the target, similar to Koriat (1993). Administration of lorazepam led to deficits in recall for the tetragrams, however no effect was found for FOK accuracy. Indeed, both placebo and lorazepam groups showed similar effects of partial information on FOK, as FOK judgements increased as more partial information was retrieved. In sum, it would appear that although benzodiazepines impair memory performance, they do not impair an individual's ability to accurately judge the contents of memory.

1.3.4.3 Neuropsychological populations

There have been a number of studies examining the effect of various neurological impairments on memory and FOK accuracy. For example, Le Berre et al. (2010) observed that patients with chronic alcoholism were impaired on episodic memory tasks and showed a likewise deficit in episodic FOK accuracy. Similarly, Schmitter-Edgecombe & Anderson (2007) observed episodic memory and FOK deficits in patients with moderate to severe closed head injury, and sufferers of mild cognitive impairment also exhibit matching episodic memory and FOK problems (Anderson & Schmitter-Edgecombe, 2010; Perrotin, Belleville, & Isingrini, 2007).

A number of disorders have also exhibited a clear dissociation between effects on episodic and semantic memory and FOK. Patients with Obsessive Compulsive Disorder (OCD) show a similar pattern of memory deficits to older adults: episodic memory is impaired and yet semantic memory is preserved (Woods, Vevea, Chambless, & Bayen, 2002). When examining FOK performance, both Jurado, Junqué, Vallejo, Salgado, & Grafman, (2002) and Tuna, Tekcan, & Topçuoğlu, (2005) found deficits not only in episodic memory but also in episodic FOK accuracy. OCD patients' gamma was at chance level in both studies, indicating that their judgements of future recognition had no predictive accuracy. In contrast, when utilising a semantic memory task, Tekcan, Topçuoğlu, & Kaya, (2007) observed no deficit in either memory or FOK accuracy for OCD patients when compared to controls. For OCD it would therefore appear that metacognitive ability when predicting future recognition mirrors recall performance. Likewise, examination of patients with Parkinson's disease follows the episodic semantic split. Baran, Tekcan, Gürvit, & Boduroglu, (2009) observed episodic memory and episodic FOK deficits, with gamma at chance level and significantly lower than control subjects. With a semantic task, Ivory, Knight, Longmore, & Caradoc-Davies, (1999) observed no memory deficit or FOK deficit.

Although the above studies appear to imply that deficits in memory performance relate to deficits in FOK accuracy, evidence from other patient populations show that the story is not quite so simple. Not only are semantic and episodic FOK dissociable, but memory performance and FOK performance are also dissociable. An observed deficit in either episodic or semantic memory is not sufficient to assume that a deficit in FOK performance will also be exhibited. Indeed, patients with temporal lobe epilepsy have been found to have impaired episodic memory and yet be at a similar accuracy level to controls when predicting future recognition performance (Howard et al., 2010). Schizophrenia patients show deficits in both semantic and episodic memory, and yet, again, both semantic (Bacon, Danion, Kauffmann-Muller, & Bruant, 2001) and episodic (Bacon & Izaute, 2009) FOK accuracy is preserved.

The picture is even more complicated in Alzheimer's disease. Here, episodic memory deficits have been found to be associated with episodic FOK deficits (Souchay, Isingrini, & Gil, 2002), and yet semantic memory deficits are not always associated with semantic FOK problems. Bäckman & Lipinska, (1993) and Lipinska & Bäckman, (1996) both found that although semantic memory was impaired in patients with Alzheimer's Disease, their semantic FOK performance was intact. However, Pappas et al., (1992) found a semantic FOK deficit in conjunction with a semantic memory deficit. It is difficult to account for these contradictory findings. It may be that patients are able to accurately monitor semantic memory depending on the subset of items they are provided with: while Pappas et al. (1992) used an existing set of questions (Nelson & Narens, 1980), Bäckman & Lipinska, (1993) and Lipinska & Bäckman, (1996) both created their own question sets. FOK accuracy has been shown to be highly dependent on the specific properties of the items sampled (Koriat, 1995), therefore if the deficit in semantic FOK accuracy is only slight, uncovering that deficit could be even more dependent on the cue questions used. It is also possible that the level of cognitive deficit between the patient groups used may have had an impact. Bäckman & Lipinska, (1993) and Lipinska & Bäckman, (1996) both describe their patients as having a mild level of cognitive impairment, based on their mean Mini Mental State Examination scores being 23.56 and 24.76 respectively. Pappas et al. (1992) judged their patients to exhibit moderate cognitive decline, based on a series of neuropsychological tests and clinical assessments. Although difficult to ascertain whether the patient groups here are comparable in terms of general cognitive ability, it may again be the case that only a slight deficit is found and, with increased cognitive impairment, the deficit in semantic FOK accuracy becomes easier to detect.

1.3.4.4 Summary

Using a variety of methods and a variety of populations, the link between memory performance prior to FOK and the accuracy of subsequent FOK judgements is complex. It may be expected that manipulations of recall accuracy, whether experimentally or with neuropsychological populations, would affect FOK in the same manner. For example, increases in recall accuracy may reflect a general increase in memory strength, thus items which are unrecalled are nonetheless closer to activation and thus are associated with a greater amount of partial information, leading to higher FOK accuracy. However, it may also be that increasing recall accuracy merely reduces the number of items on which the FOK can be measured, leading to greater variability and affecting measurement of accuracy, or the amount of partial information available for unrecalled items remains similarly distributed irrespective of recall performance, thus no change in

FOK accuracy is found. FOK judgements are based on heuristics and indirect assessments of partially retrieved information about the unrecalled target rather than direct access to the target memory trace, therefore FOK judgements are unlikely to ever be completely accurate in their predictions. Nonetheless, understanding where associations between memory performance and FOK accuracy do and do not occur can provide insight into the cognitive process involved as well as enabling more complete descriptive and diagnostic tools within clinical settings.

1.3.5 Beyond the laboratory: Real world applications of FOK and metamemory

So far the contributions of FOK research have been considered in experimental settings with standard laboratory paradigms. However, retrieval failures also occur in everyday memory situations, which may lead to FOK sensations and subsequently drive behaviour. For example, a song may be playing on the radio and, although you cannot name the artist currently, you may nonetheless feel as though you know who it is and would recognise the name once told. Thus it is useful to consider the FOK within more applied settings. To some extent this can be achieved merely by using stimuli which mimic naturally occurring memory tasks. For example, Hosey et al. (2009) asked participants to learn name-face pairings, finding that participants were more likely to report using cue familiarity than target accessibility when making FOKs for unrecalled names although no effect of strategy use was observed on FOK accuracy. Although the authors focused on the theoretical implications of the findings, the results also show that participants were able to accurately judge their recognition of unrecalled names when cued with faces. This ability to accurately predict memory for cross-domain associations has important implications within eyewitness memory.

The reliability of eyewitness testimony has, understandably, received considerable attention, establishing various factors and situations which can influence accurate recall and identification of suspects (Clark & Godfrey, 2009; Wells & Olson, 2003).

The relationship between the accuracy of the account provided by an eyewitness and their confidence in the veracity of this account is also of key importance. People may be more easily swayed by an eyewitness who confidently asserts incorrect details of a crime than one who less confidently reports correct details (Keren & Teigen, 2001; Price & Stone, 2004; Thomas & McFadyen, 1995). Thus an eyewitness who is confident they will recognise an unrecalled detail may be believed more than an eyewitness who is not confident they would recognise an unrecalled detail, irrespective of subsequent recognition accuracy.

Two key studies have considered the accuracy of FOK judgements in eyewitness situations. Perfect & Hollins (1996) contrasted FOK accuracy for an eyewitness task to that for general knowledge questions. Participants were presented with a short film clip to simulate an eyewitness event. One day after encoding, participants were asked a series of 50 questions regarding the video clip, and FOK judgements were elicited for unrecalled items. In addition, 50 general knowledge questions were also given, again with FOKs obtained for unrecalled items. Recall accuracy for the two tasks was similar, suggesting a similar level of difficulty between the two memory tasks. However, FOK accuracy was not equivalent. While participants were able to predict future recognition within the general knowledge task, accuracy for the eyewitness task was at chance level and significantly below that of the general knowledge accuracy. Participants were unable to identify those items which would later be recognised from those which would not. This lack of predictive accuracy was replicated by the same authors within a series of three experiments, all of which observed FOK performance to be at chance level (Perfect & Hollins, 1999). Thus although an eyewitness may feel confident that they will subsequently recognise unrecalled details about a crime, they are unlikely to actually be able to identify the correct details when prompted.

Findings from the metacognitive literature can also be applied to education. A positive feeling of knowing judgement gives the impression that the information is known but cannot currently be accessed. It could be considered a feeling of encoding, that the information has been sufficiently learned to enable retrieval, there is simply a temporary blockage to retrieval. However, it could also be the case that the sensation is erroneous, that the information cannot be retrieved as it has not been sufficiently encoded, thus cannot be accessed at a later date. For students, it is important to be aware that the FOK may or may not be a reliable indicator of their memory state. During revision for a test, for example, the student may decide that the FOK is strong enough that they need only briefly consult their notes and then continue to the next topic. However, if the FOK is strong but the encoding of the information is weak, it may be more appropriate for the student to spend more time consulting their notes and consolidating the unrecalled information.

Glenberg, Wilkinson, & Epstein (1982) considered this "illusion of knowing" in relation to text comprehension. Participants were asked to read three texts, all of which were either a single paragraph in length or three paragraphs in length. Upon reading each passage, participants rated how well they had understood the content and were also asked two true-false questions. Importantly, each passage contained a pair of contradictory sentences which participants were told to look out for. Illusions of knowing occurred when ratings of comprehension were high yet no detection of the contradiction was reported. It was found that, when presented with a single paragraph of text, the illusion of knowing rate was 14.9%. In contrast, when three paragraphs were given, this increased to 51.1%. Thus errors were not detected in over half of the longer passages which received higher ratings of comprehension, despite explicit instructions to be aware of contradictions within the text. This has clear implications in a classroom setting, where students are frequently asked to read and digest passages of much greater length than three paragraphs. Further investigation by Kroll & Ford (1992) clarified that the prevalence of illusions of knowing could be mediated by the participants motivational orientation. Those who were more task-directed were able to detect more errors and showed lower incidence of illusions than those who were less task-directed.

1.4 Aims of the thesis

The thesis presented here aims to further explore the nature of the relationship between memory performance and subsequent sensations associated with retrieval failure. This aim is achieved by manipulating memory accuracy at encoding, retention and retrieval, and by considering different populations. The primary measure of memory experience used is the feeling of knowing. Both magnitude and accuracy of feeling of knowing judgements will be examined. The magnitude of the FOK gives an indication of the bias in responding, the decision rule used to give the FOK response. FOK accuracy reflects the ability of the participant to discriminate between items which will be recognised and those which will not. Bias and accuracy may be differentially affected by memory manipulations; if a decision rule is consistently applied then no impact will occur on FOK accuracy, whereas if participants do not consistently apply the decision rule changes in FOK accuracy can occur. Thus it is important to examine both measures to ascertain where the effect of the memory manipulation is occurring: changes to bias or changes to accuracy. Chapter 2 considers the effect of healthy ageing on FOK accuracy for semantic and episodic information. In addition to considering two types of memory task, encoding was also manipulated through the use of repeated learning trials. Retention and retrieval are considered in Chapter 3, also within a healthy ageing population in addition to young adults. Chapter 4 explores the application of signal detection theory to the FOK literature, and proposes a preliminary model of FOK within this framework. Chapter 5 considers a novel measure of memory experience, related to the FOK. While the FOK is a prediction of future retrieval of unrecalled information, the measure developed here is a post-diction regarding the amount of unrecalled information. This measure was explored within healthy young and older adults, and also within patients diagnosed with dementia. Chapter 6 summarises the research findings of Chapters 2 to 5 and considers future objectives from these foundations.

Chapter 2. Semantic and Episodic Feeling of Knowing Accuracy in Ageing

2.1 General introduction

The distinction between episodic and semantic memory was first proposed by Tulving (1972). Semantic memory is a system which stores accumulated knowledge about the world. This includes definitions of words, facts such as the capital city of France, and also more complex information about how society works such as what you should expect to happen when you go to a restaurant. It is a general type of memory, one that is typically acquired over repeated exposures and so is not linked to a specific time or a specific episode. In contrast, episodic memory is very definitive. Episodic memory allows the rememberer to encode, store and retrieve single unique events or episodes that are linked to a specific time and place. So while the question ‘What did you eat for breakfast today?’ would require retrieval from episodic memory, the question ‘What do people typically eat for breakfast?’ would require retrieval from semantic memory.

Age has differential effects on each of these types of memory. For tasks measuring semantic memory (e.g. verbal fluency, general knowledge tasks), older adults are typically able to perform at similar levels to young adults. However for tasks measuring episodic memory (e.g. word list learning), older adults are not able to match younger adult performance; problems at encoding and/or retrieval lead to lower levels of accuracy in older adult samples (see Anderson & Craik, 2000; Zacks, Hasher, & Li, 2000 for reviews).

Although the effect of age on semantic and episodic memory is well documented, the effect on monitoring and control processes associated with each type of memory is less clear. As discussed in Chapter 1, the feeling of knowing (FOK) paradigm can be used to examine peoples’ ability to predict future recognition of unrecalled items for both semantic and episodic items, and so is a useful tool to examine age effects on both types of memory. In terms of semantic FOK, results are consistent with the memory literature in that no age effect is observed. Butterfield, Nelson, & Peck (1988), Lachman, Lachman, & Thronesbery (1979), Bäckman & Karlsson (1985) and Marquié & Huet (2000) all used a general knowledge task similar to that of the original Hart (1965) task. Whether using a binary FOK (Butterfield et al., 1988), a rating scale FOK (Lachman et al., 1979; Marquié & Huet, 2000) or relative FOK judgements (Butterfield et al., 1988) no differences were observed in the predictive accuracy of young and older adults judgements for future recognition. Likewise, Allen-Burge & Storandt (2000) examined semantic memory and FOKs for rare word definitions, again finding similar levels of predictive accuracy of FOK judgements in young and older adults. The preservation of semantic memory processes in ageing would appear to extend to memory monitoring processes of semantic memory.

When considering episodic FOK accuracy, the effect of ageing is an issue of some debate. In agreement with the findings of the memory literature, Perrotin, Isingrini, Souchay, Clarys, & Taconnat (2006), Souchay, Isingrini, & Espagnet (2000), and Thomas, Bulevich, & Dubois (2011) have all observed deficits in episodic FOK accuracy in older adults for word pair learning, whether using binary FOKs or ratings (the implications of using binary or ratings based FOK judgements within a task will be considered in Chapter 4). So, when learning new material, it would appear that older adults are unable to accurately monitor the contents of their memory, therefore are unable to predict future recognition of unrecalled items. However, Hertzog, Dunlosky, & Sinclair (2010) and MacLaverty & Hertzog (2009) failed to find an episodic FOK deficit. Although a similar word pair learning paradigm was used, older adults in these two studies were able to predict their future recognition with a similar level of accuracy to that of young adult participants. The question of why there is a lack of agreement within the literature is addressed here.

First we must consider why there may potentially be an age effect on episodic FOKs but not semantic FOKs. As explained in Chapter 1, the link between memory and

FOK performance is not always straightforward. Therefore the observation of an age effect on memory is not in itself sufficient to indicate a similar deficit in FOK accuracy. Three lines of evidence will therefore be presented which indicate that the assumption of an episodic impairment in healthy ageing is a reasonable one: neuropsychology, executive function and recollection.

Neuropsychology

There have been a number of studies examining the effect of various neurological impairments on memory and FOK accuracy (see Chapter 1 for a more in depth review). Importantly, these studies have shown that a dissociation can occur between semantic FOK and episodic FOK accuracy, suggesting monitoring of these two types of memory may rely upon different cognitive processes and/or physical structures. The clearest examples of this dissociation occur in patients with Obsessive Compulsive Disorder (OCD) and patients with Parkinson's disease. Tekcan, Topçuoğlu, & Kaya (2007) observed preserved semantic memory and semantic FOK accuracy in patients with OCD. In contrast, Jurado, Junqué, Vallejo, Salgado, & Grafman (2002) and Tuna, Tekcan, & Topçuoğlu (2005) both found deficits in episodic memory accuracy and subsequently in episodic FOK accuracy. The same pattern of results was also shown in Parkinson's disease patients, with semantic memory and semantic FOK preserved (Ivory et al., 1999) while episodic memory and episodic FOK is impaired (Baran et al., 2009).

The pattern of deficits within these patient groups mirrors that predicted in healthy ageing. However, it is important to bear in mind that deficits in memory accuracy do not always relate to deficits in FOK accuracy. For example, patients with schizophrenia exhibit deficit sin both semantic and episodic memory tasks, and yet no evidence has been found of a deficit in semantic or episodic FOK accuracy (Bacon et al., 2001; Bacon & Izaute, 2009). Thus although evidence from neuropsychological studies is encouraging for the assumption of a selective deficit in episodic FOK accuracy in ageing, it is not in itself sufficient to support this assumption.

Executive function

Declines in various aspects of cognitive performance in older adults have been reported in areas including speed of processing (Salthouse, 2000), attention (Kok, 2000; Plude, Enns, & Brodeur, 1994) and executive function (Goh, Beason-held, An, Kraut, & Resnick, 2013; Treitz, Heyder, & Daum, 2007) as well as in episodic memory tasks (Anderson & Craik, 2000; Zacks et al., 2000). Efficient cognitive processes rely on intact brain structures and brain chemistry. Neuroimaging has allowed examination of the exact changes that occur to brain anatomy in normal ageing, and how these impact performance on various cognitive tasks. Although decline occurs throughout all brain structures (Raz, 2000), some areas are more susceptible than others. In particular, the frontal lobes are especially sensitive to the effects of increased age, showing the earliest signs of atrophy and also the greatest level of deterioration (Raz et al., 2005; Dennis & Cabeza, 2008). This increased susceptibility of certain brain regions to deterioration ageing could lead to differential effects on semantic and episodic FOK accuracy, as evidence suggests that as well as sharing a network of regions these two processes also recruit different brain structures (e.g. Reggev, Zuckerman, & Maril, 2011; see Chapter 1).

Although no direct comparison of patterns of brain activation in ageing during semantic and episodic FOK tasks has yet been carried out, it is possible to examine the behavioural effects of age-associated decline in brain function and their relationship to FOK tasks. Fernandez-Duque, Baird, & Posner (2000) noted that metacognition is closely related to executive function, and executive function is highly reliant on frontal lobe integrity (Smith & Jonides, 1999; Stuss, 2011). Indeed, the importance of this brain region in making episodic FOK judgements has been demonstrated by work with frontal lobe patients, who are unable to make accurate judgements of future recognition for episodic information (Janowsky, Shimamura & Squire, 1989; Pinon, Allain, Kefi, Dubas, & Le Gall, 2005; Schnyer, Nicholls, & Verfaellie, 2005; but see Pannu, Kaszniak & Rapcsak, 2005). Both metacognition and executive function involve the ability to monitor and control the processing of information.

Although these two areas of research have largely remained separate, a number of studies have now examined the relationship between FOK accuracy and performance on tasks measuring executive function, particularly in ageing. Souchay, Isingrini, & Espagnet (2000) used two measures of executive function typically administered in neuropsychological examinations: the Wisconsin Card Sorting Task (WCST) and the Verbal Fluency Test (FAS). As expected, older adults showed considerable deficits on the two measures of executive function. In addition, both young and older adults' episodic FOK performance was correlated with the executive function scores, supporting the hypothesis that these two cognitive processes are related. Finally, by partialing out the measures of frontal lobe functioning, age-related variance in older adults' FOK accuracy was reduced by 86%, confirming that executive function has a key role to play in the ability to accurately predict future recognition in the FOK task. Further studies have confirmed executive functioning as a mediator for episodic FOK performance in ageing (Perrotin et al., 2006; Souchay, Isingrini, Clarys, Taconnat, & Eustache, 2004; Souchay & Isingrini, 2004).

Recollection

A final line of evidence that would also suggest an age-related episodic FOK deficit is recollection. As mentioned previously, Tulving (1972) proposed the distinction between semantic and episodic memory. As well as relying on different memory stores, the subjective states experienced during retrieval also differ between the two types of memory (Tulving, 1985b). During semantic memory retrieval, only the memory itself is accessible. This leads to the sensation of 'knowing', the memory is retrieved in isolation therefore only the factual details of the memory itself are accessible. In contrast, retrieval from episodic memory typically involves extra details as well as the memory itself. The images, thoughts and feelings that are accessible in addition to the memory lead to the sensation of 'remembering', a much more enriched type of retrieval, often referred to as 'mental time travel' (Tulving, 1985b). One way in which to measure recollection is to use the Remember/Know procedure, whereby participants are asked to categorise their subjective experience during retrieval as either ‘remembering’ or ‘knowing’ (see Gardiner, Ramponi, & Richardson-Klavehn, 2002 for a review).

Research using this paradigm has shown that in healthy ageing, participants show a decrease in remembering (Bugaiska et al., 2007; Bunce, 2003; Clarys, Bugaiska, Tapia, & Baudouin, 2009; Friedman, de Chastelaine, Nessler, & Malcolm, 2010; Perfect & Dasgupta, 1997; Souchay et al., 2007). This would suggest that older adults are unable to retrieve the extra detail associated with the memory that leads to the experience of recollection. In addition to the subjective measure of Remember/Know judgements, objective measures also indicate that ageing is associated with a deficit in recollective experience. Older adults show difficulties in tasks involving source memory (Dennis et al., 2008; Henkel, Johnson, & De Leonardis, 1998; Mitchell & Johnson, 2009; Thomas et al., 2011), whereby they are unable to report details which occurred simultaneously with target presentation during encoding. The process dissociation procedure (Jacoby, 1991) likewise has shown that older adults exhibit deficits in recollection, while familiarity processes remain intact (Benjamin & Craik, 2001; Caldwell & Masson, 2001; Jennings & Jacoby, 1993).

The link between FOK and recollection was first explored by Hicks & Marsh (2002). In addition to making FOK judgements about unrecalled items, during the recognition task participants were also asked to make Remember/Know judgements about their retrieval experience. These two judgements were then analysed, with the finding that higher FOK ratings were associated with more Remember responses on the subsequent recognition task. Souchay et al. (2007) suggested that the partial cues and inferential mechanisms activated during a failed recall attempt may not only be used to establish how likely future recognition of a missing item may be, but also as a basis for recollective experience. As Koriat (1993) proposed, it is the amount of partial information retrieved about an inaccessible target that drives the feeling of knowing judgement. This partial information, which may include contextual details such as feelings and thoughts which occurred during encoding, would therefore give the extra details necessary to elicit the sensation of recollection rather than familiarity. Arguably, recollection and FOK both rely to some extent on the same information being processed. As there is a well documented age-related decline in recollective experience in ageing, it would therefore seem reasonable to assume that there would also be age-related declines in episodic FOK accuracy. It is this specific association between recollection and FOK in older adults that was examined in Experiment 2 of Souchay et al. (2007). The same young and older participant groups took part in both a standard episodic FOK task and a standard Remember/Know task. Accuracy in the FOK task (as measured by Gamma) showed a clear relationship with the proportion of Remember responses, but not with Know responses. Hierarchical regression further showed that the age effect observed in the FOK task was primarily explained by the age effect on recollection.

Investigations into source monitoring also shed light on the link between feeling of knowing and recollection. Judgements about the source of the information retrieved require access to additional details in conjunction with the target itself, and are therefore thought to be indicative of recollection (Guttentag & Carroll, 1997; Perfect, Mayes, Downes, & Eijk, 1996; Yonelinas, 1999; but see Hicks, Marsh, & Ritschel, 2002). In ageing, older adults fail to use source information details to boost recollection (Skinner & Fernandes, 2009). Thomas et al. (2011) looked at the effect of source retrieval on FOK judgements. Over 2 experiments it was found that older adults did not spontaneously use retrieved source details to inform their FOK judgements, leading to poorer accuracy than young adults. However, if explicit instructions were given to utilise the source information when judging future recognition, older adults were able to improve their FOK accuracy, although they still exhibited some deficit compared to young adults.

Despite evidence indicating a selective impairment in episodic FOK accuracy might be expected in ageing, some studies have found evidence for preservation of accuracy (Hertzog et al., 2010; MacLaverty & Hertzog, 2009). The memory constraint hypothesis (MCH; Hertzog et al., 2010) proposes that FOK accuracy is primarily dependent on the quality of the underlying memory processes. If encoding is impaired in some way, then insufficient or incorrect partial information will be accessed during the failed recall attempt. Subsequently, the FOK judgement will be based on these flawed details, leading to lower predictive accuracy (Koriat, 1993, 1997). As episodic memory is impaired in ageing, the FOK deficit observed in some studies may simply be due to a lack of sufficient encoding to allow diagnostic partial information to be available to the participant as opposed to a deficit in metacognitive ability (see Perfect & Stollery, 1993 for a similar argument). Indeed, by equating young and older adults' memory performance using variable delays, Hertzog et al. (2010) were able to demonstrate equivalent levels of episodic FOK accuracy in the two age groups. However, although some of the studies which have found age effects on FOK do also show age effects on memory performance (Perrotin et al., 2006; Souchay et al., 2000, 2007), not all of them do. For all three experiments reported by Thomas et al. (2011), memory performance on the cued recall aspect of the task was comparable between the young and older adult participants. Despite this, older adults consistently demonstrated a deficit in episodic FOK accuracy. These findings do not support the MCH explanation of the relationship between ageing and FOK, and so suggest that the relationship is far more complex than the MCH assumes.

In sum, semantic FOK accuracy appears to be preserved in ageing, whereas for episodic FOK accuracy the picture is less clear. Despite evidence from related literature (e.g. brain imaging, patient work, recollection) to suggest a selective episodic FOK impairment would be of no surprise in an older adult population, there is also evidence of preserved episodic FOK accuracy that must be taken into account. The purpose of the current chapter is to establish whether, by using the same participants and the same task, a selective deficit in episodic FOK performance can be detected. Experiment 2.1 uses a general knowledge paradigm as has been previously used within the FOK literature to explore semantic FOK. Items which were not recalled at the semantic test were then used to test episodic FOK. Experiment 2.2 uses a novel language learning paradigm, and again items not recalled at the semantic test were used to examine episodic FOK.

2.2 Experiment 2.1: General knowledge task

2.2.1 Introduction

When considering whether age does lead to a selective impairment in episodic FOK accuracy it is important to consider one key limitation of the studies discussed so far. Examination of semantic and episodic[1] FOK accuracy has typically been conducted in isolation, therefore the accuracy of each of these processes has been established in different participant samples and then compared. To date, only one study has examined both semantic and episodic FOK accuracy within the same group of participants. Souchay et al. (2007) asked participants to complete both a general knowledge task and a word pair learning task, thereby allowing them to directly compare accuracy in the two tasks. In addition, the same target items were used in each task, allowing a further level of control in the study. For example, in the semantic task the question may have been 'What was the subject of Magritte's famous surrealist painting La Trahison?', the answer being 'Pipe'. In the episodic task, an unrelated cue word would be paired with the same item, e.g. Birthday – Pipe. The intrinsic properties of target words, such as frequency, can have an impact on the FOK judgement (Koriat, 1993). By using the same target items, and by counterbalancing task order, Souchay et al. (2007) were able to match the properties of the required target in the semantic and episodic tasks. For the semantic task, as in previous research, comparable performance was observed between young and older adults: both groups were able to accurately predict future recognition accuracy for unrecalled items. When examining performance on the episodic task, both young and older adults' gamma scores were above chance, indicating that both age groups were able to predict their performance to a certain extent. However, the accuracy of the older adult group was shown to be significantly below that of the young adult group. Despite exhibiting some ability to judge their recognition of unrecalled items, older adults were unable to do this to

[1] For the purposes of this chapter, the terms 'episodic' and 'semantic' refer to the general definition of these processes, as described in the general introduction. Where the terms 'Episodic' and 'Semantic' are used, we are referring to their conceptualisation for the current paradigm, as defined in the Method sections.

the same level of accuracy as young adults. This would therefore support previous findings observing a selective episodic FOK deficit in older adults (Perrotin et al., 2006; Souchay et al., 2000; Thomas et al., 2011).

Although Souchay et al. (2007) ensured the target items were identical in the semantic and episodic task, thus removing target characteristics as a potential confound on FOK accuracy, task characteristics could still be a factor. The prompt used to elicit the recall attempt in each task is different, with the semantic task involving a more conceptual cue whereas the episodic task involves a more contextual cue (Koriat, 1997; Perfect & Hollins, 1999). This may influence the strategies used at the recall attempt, thereby helping to determine the quality and quantity of partial information accessed on which the FOK judgement will be based. The present study therefore aims to remove this possible influence of cue characteristics by utilising a general knowledge task for both the semantic and episodic FOK tasks. Thus, prior to learning, the general knowledge items tap semantic memory. Items which are not known at this stage will subsequently be learned, therefore the unknown general knowledge questions become the episodic version of the task. The critical feature of the task must be that sufficient items are remaining following the semantic recall attempt. However, the task must not be of too high a difficulty level overall, otherwise participants may become discouraged with their performance and not fully attempt to retrieve the answers. A general knowledge task is therefore highly suitable as it is possible to manipulate the difficulty of individual items. It can be ensured that sufficient items are included to motivate participants to maintain retrieval effort throughout the task, but also include items of higher difficulty level that would require learning in order to answer correctly.

In addition to measuring semantic and episodic FOK judgements, Remember/Know judgements were also obtained during the episodic recognition test. Episodic FOK and recollection have been proposed to rely on similar access to partial information (Souchay et al., 2007). This relationship between the two subjective experiences, combined with an observed age deficit in recollection, has been used to infer a selective age deficit in episodic FOK accuracy. However, although Souchay et al., (2007) measured both processes within the same young and older adult participants, different tasks were used. Therefore cue characteristics could have had an influence. The present study measured episodic FOK and Remember/Know judgements for both young and older adults for the same target items, as undertaken previously by Hicks & Marsh (2002) with young adults only.

2.2.2 Method

2.2.2.1 Participants

Thirty five undergraduate students (age range 18 to 29, *M* = 20.23, *SD* = 2.89; nine male) from the University of Leeds participated in the study in return for course credit. The older adult group consisted of 21 people aged between 60 and 85 (*M* = 69.86, *SD* = 8.00; three male) recruited from the local community. All older adults obtained scores on the Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) above the cut off of 27 (*M* = 28.79, *SD* = 1.44) and were not taking medication which would affect cognitive function.

2.2.2.2 Materials

An initial set of 50 general knowledge questions was created and administered to five young adults as pilot participants who did not take part in the main study. This allowed assessment of the ease of the questions and the number of items that would elicit FOK responses i.e. errors of omission. Average correct recall was 11 items, with 23.4 items leading to an FOK judgement. Recognition accuracy was at 19.4 items of the 50 questions presented. Based on this, the number of items for the final study was increased to allow greater scope for learning for the Episodic aspect of the task. The final materials consisted of 60 questions covering a variety of topics including sport, history, geography and general knowledge. Questions were printed in two answer booklets, one for Semantic recall and recognition and one for Episodic recall and recognition. Calibri font (font size 11) was used throughout the booklet for presentation of recall prompts, recognition test and learning of correct responses. For the recall test, the question was presented together with a box for participants to write the answer if known. Binary FOK judgements were elicited by the question *'Would you recognise the correct answer if it was given to you?'* presented beneath the response box with a Yes/No option. Participants were asked not to guess when completing the recall task, and to give FOK judgements only for items where they were unable to provide an answer. A four-alternative forced choice recognition task was used, with distracters being plausible answers to the question presented. For example, to the question '*Which channel island is nearest to France?'* the names of three channel islands, *Guernsey*, *Jersey* and *Sark,* were given as distracters in addition to the correct answer of *Alderney*. Question order was randomised for each recall and recognition test for the Semantic and Episodic tasks. A second set of booklets was created to allow counterbalancing of the question order.

2.2.2.3 Procedure

All participants were tested individually. A standard recall-judgement-recognition procedure was used to obtain FOKs, similar to those previously utilised in Semantic (Bäckman & Karlsson, 1985) and Episodic (Schacter, 1983) FOK tasks. The key difference in the administration of Semantic and Episodic FOK tasks is the inclusion of a learning stage prior to recall in the Episodic FOK task only. As noted in the introduction, the novelty of the current procedure is to capture Semantic and Episodic FOK measures from the same participants for the same target items which have also been elicited by the same cues i.e. the memory search for the target item '*Alderney'* is prompted by the question '*Which channel island is nearest to France?'* for both the Semantic FOK and the Episodic FOK. Due to this procedural necessity, all participants completed the Semantic task followed by the Episodic task. Participants were provided with written instructions for the task, and two example questions were provided to ensure participants understood the FOK judgement. Questions were presented in a different order at each test of recall and recognition and also at learning.

Semantic task. The procedure included a cued-recall phase, a FOK judgement phase, and a recognition phase. All recall, FOK and recognition responses were recorded by participants in the booklet provided. In the recall test, the 60 questions were presented with a response box beneath. Participants were asked to write the correct answer if known, or to leave the response box blank if the correct answer was not known. Guessing was discouraged. Immediately after the recall attempt, the FOK judgement was made. FOK judgements were only made for those items where a response was not attempted i.e. omission errors. The FOK was prompted with the question *'Would you recognise the correct answer if it was given to you?'.* If participants thought they would be able to recognise the correct answer, they circled a Yes response; if participants thought that they would not recognise the correct answer, they circled a No response. Once recall or FOK had been completed for the question, participants then moved on to the next question until all 60 questions had received a response. The recognition phase followed immediately after. The 60 questions were presented again with four alternative responses, the correct response plus three plausible distracters. Participants were asked to circle the correct answer to each question, and to guess where they were unsure of the answer, until all 60 questions had been assigned an answer. No time limit was imposed for any phase of the Semantic task.

Episodic task. The procedure included a study phase, a cued recall phase, a FOK phase and a recognition phase. In the study phase, participants were presented with the same 60 questions as those in the Semantic task together with the correct responses. All 60 questions were provided to avoid giving feedback to participants as to the accuracy of their responses during the Semantic task. Questions were printed on two sheets of paper with the correct answer highlighted in bold. Presentation was timed at 120 seconds, and participants were asked not to mark the sheets in any way. A response booklet was then provided to record responses. Cued recall and FOK judgements were obtained in the same way as in the Semantic task, with the 60 questions printed in the booklet with recall response boxes and FOK prompts, and the instruction not to guess. Upon completion of either a recall or FOK response for all questions, the recognition phase was completed. This again took the form of a four-alternative forced choice, with the same distracters used as in the Semantic task. In addition, participants were asked to report their recollective experience. After selecting an answer, participants were also asked to categorise that response as Remember, Know or Guess. The following definitions (taken from Conway, Gardiner, Perfect, Anderson, & Cohen, 1997) were given:

> A REMEMBER response would be when you can remember a specific episode from when you were learning the correct answer. You might have had specific thoughts or feelings regarding the answer, or a specific memory may have come to mind when viewing the question and answer.
>
> A KNOW response is when you "just know" the correct answer and the alternative you have selected "stood out" from the four choices available. In this case you would not recall a specific episode and instead you would simply know the answer.
>
> A GUESS response is when you have not remembered or known the answer. In this case you may have made a guess, possibly an informed guess, e.g. some of the choices look unlikely for other reasons so you have selected the one that looks least unlikely.

These definitions were provided prior to the recognition phase and were available throughout recognition for participants to refer to.

2.2.3 Results

Consideration of which items to class as Episodic needs clarification. All items were included in the analysis of Semantic memory and metacognitive performance. However, during the Episodic task procedure, all items were presented for a second time, thereby including items which were already stored as semantic knowledge. This would therefore contaminate the Episodic measure. The analysis presented below considers two classifications of Episodic items. First, a liberal criterion of what is Episodic was used. In this instance, all items which were correctly recalled during the Semantic task were removed from the Episodic analysis. The adoption of a liberal criterion does have associated issues to be aware of. It is possible that an item which is not correctly recalled may nonetheless have a semantic representation, and thus should not be included in the assessment of

Episodic FOK accuracy. Therefore the second analysis employs a strict criterion based on recognition performance, whereby items which were correctly recognised during the Semantic task were subsequently removed from the Episodic analysis. This does employ some amount of conservatism as not all items may have been correctly recognised due to access to a memory, participants may have simply guessed the correct response. Nevertheless, this strict criterion ensures contamination between Semantic and Episodic tasks is minimised. The data was therefore analysed two ways. All participants provided the same number of data points during the Semantic task, while for the Episodic task each participant had a varying number of Episodic items based on their performance in the Semantic memory measures. Therefore all data is presented as proportion accuracy.

2.2.3.1 Liberal criterion

Four older adult participants were removed from the analysis due to no errors of omission occurring i.e. responses were given for all questions on the Semantic trial. With no errors of omission, no items could be used to establish Episodic FOKs. The subsequent analysis involved 35 young adults (age *M*= 20.23, *SD* = 2.89) and 17 older adults (age *M* = 68.65, *SD* = 7.95). For the Semantic task, all participants responded to all 60 questions, thus responses are divided by 60 to give proportion measures. For the Episodic task, responses are divided by the number of trial each participant failed to recall correctly at the Semantic task, thus proportion calculations are calculated individually. A mean number of 50.173 (*SD* = 7.334) trials comprised the Episodic task.

Memory

To assess recall, the number of items correctly recalled was divided by the total number of possible items (i.e. 60 for the Semantic task, individually calculated for the Episodic task) to give the proportion correct. Likewise, recognition accuracy was calculated by dividing the number of items correctly responded to by the total number of possible items. Mean proportions of correct recall and recognition accuracy for each age group are shown in Figure 2.1. A 2 (age: young adults and older adults) x 2 (task: Semantic and Episodic) repeated measures ANOVA was conducted on recall performance. A main effect of age was observed, $F(1,50) = 25.559$, $p < .001$, $\eta^2 = .338$, with older adults showing higher recall accuracy than young adults. A main effect of task was also present, $F(1,50) = 407.169$, $p < .001$, $\eta^2 = .891$, indicating that Episodic recall performance was greater than that of Semantic recall. An interaction between age and task was found, $F(1,50) = 5.562$, $p = .022$, $\eta^2 = .100$. Follow up t-tests indicated that although both groups showed higher accuracy in the Episodic compared to the Semantic task, this difference was greater in young adults than for the older adults (young adults: $t(34) = 20.733$, $p < .001$; older adults: $t(16) = 9.894$, $p < .001$).

Recognition accuracy was also examined with a 2 (age) x 2 (task) ANOVA, showing higher accuracy in the older adult group compared to the young adults, $F(1,50) = 16.845$, $p < .001$, $\eta^2 = .252$. Recognition was higher in the Episodic task than the Semantic task, $F(1,50) = 384.607$, $p < .001$, $\eta^2 = .885$. An interaction between age and task was also present, $F(1,50) = 42.751$, $p < .001$, $\eta^2 = .461$, with young adults again showing a greater difference in performance in the Episodic task compared to the Semantic task than the older adults (young adults: $t(34) = 23.833$, $p < .001$; older adults: $t(16) = 7.376$, $p < .001$). This difference in accuracy shown by the young adults does in fact remove the effect of age, with recognition in the Semantic task showing a clear age effect, $t(50) = 7.272$, $p < .001$, whereas for the Episodic task no age effect remains, $t(50) = 0.003$, $p = .998$.

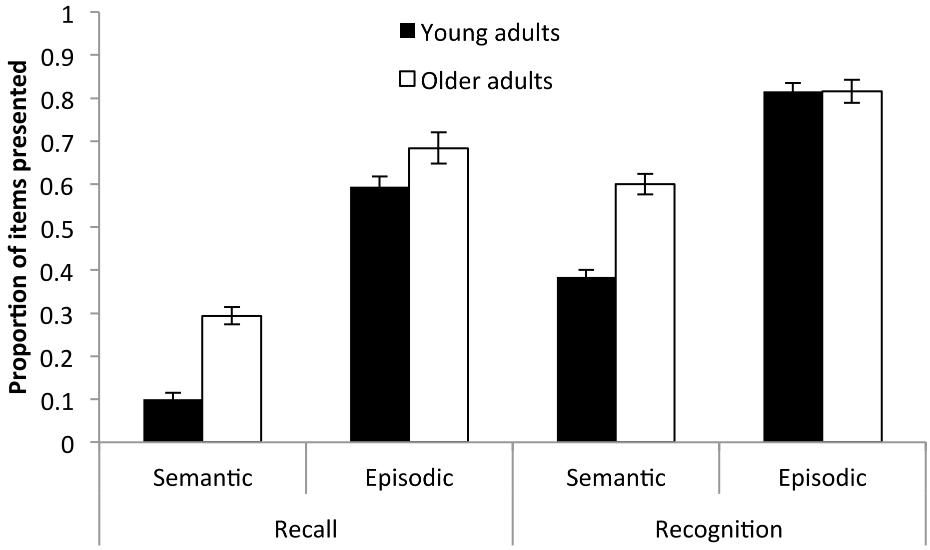


Figure 2.1: Proportion of total items presented which were correctly recalled and recognised for young and older adults in each condition based on the liberal criterion

In addition to correct recall, the proportion of items where no recall attempt was made (i.e. errors of omission) is of interest. It is these items which will comprise the FOK analysis. Errors of commission may also indicate whether one group is more likely to attempt a recall response, indicating a liberal criterion for explicit retrieval. A 2 (age) x 2 (task) repeated measures ANOVA revealed a main effect of age for errors of omission, $F(1,50) = 32.068$, $p < .001$, $\eta^2 = .391$, with young adults providing more errors of omission than older adults (see Table 2.1). More errors of omission were also given in the semantic task than the episodic task, $F(1,50) = 433.737$, $p < .001$, $\eta^2 = .897$. An interaction between group and task was also present, $F(1,50) = 9.518$, $p = .003$, $\eta^2 = .160$, due to a greater disparity in the proportion of errors at the semantic task than the episodic task between the two age groups (follow up independent t-tests: Semantic task $t(50) = 6.429$, $p < .001$, Episodic task $t(50) = 3.088$, $p = .003$; follow up paired samples t-tests: young adults $t(34) = 24.846$, $p < .001$, older adults $t(16) = 8.495$, $p < .001$). Errors of commission also show a main effect of group, $F(1,50) = 9.790$, $p = .003$, $\eta^2 = .164$, with older adults giving more errors of commission than young adults. Thus older adults would appear to have a lower recall criterion than young adults, although levels of commission errors are quite low for both age groups. No effect of task, $F(1,50) = 1.013$, $p = .319$. $\eta^2 = .020$, or interaction between age and task, $F < 1$, was found.

Table 2.1: Mean proportions of errors of omission and errors of commission during recall for Semantic and Episodic tasks in young and older adults

		Young Adults	Older Adults
		M (SD)	**M (SD)**
Omission errors	Semantic	0.845 (0.093)	0.569 (0.165)
	Episodic (liberal)	0.360 (0.155)	0.208 (0.187)
	Episodic (strict)	0.386 (0.132)	0.227 (0.179)
Commission errors	Semantic	0.054 (0.010)	0.137 (0.015)
	Episodic (liberal)	0.047 (0.021)	0.108 (0.031)
	Episodic (strict)	0.061 (0.059)	0.141 (0.104)

The pattern of errors obtained for recall suggests that, although the proportion of incorrect responses given is low, older adults are more likely to give an incorrect response than younger adults. Additionally, older adults give fewer errors of omission than younger adults, meaning that subsequent FOK analysis is based on a smaller number of responses, and potentially a smaller range of possible FOK sensitivity. It is interesting to note that the typical effect of age on memory is not present. Rather than older adults showing lower Episodic memory performance, for recall older adults are in fact significantly more accurate than young adults, although this age advantage does disappear at recognition. Nonetheless this provides an unexpected further test of the MCH.

Metacognition

It is important to note that the FOK is based on unrecalled items only. Items which are incorrectly recalled are not included in the FOK analysis, as these items have a different relationship to future recognition than items where no response is given (Krinsky & Nelson, 1985). However, sufficient items are needed to ensure stability in the assessment of FOK accuracy. As fewer items are available for the Episodic task due to the design used, the number of FOK responses available was considered before analysis of proportional data. For the Semantic task, young adults gave an average of 50.686 (*SD* = 5.630) FOK responses, while older adults gave 34.059 (*SD* = 9.865) responses. For the Episodic task, the number of FOK judgements made is much lower: an average of 19.257 (*SD* = 7.872) for young adults and 9.412 (*SD* = 9.988) for older adults. The low number of FOK judgements given in the Episodic task by older adults needs to be taken into consideration when interpreting further analyses.

To examine the FOK, the proportion of Yes FOK responses given were analysed first to establish whether a bias in responding occurred. An indication of FOK accuracy was then considered by comparing the proportion of Yes FOKs which were subsequently recognised to the proportion of No FOKs which were subsequently recognised. FOK accuracy is indicated if recognition of Yes FOK items exceeds recognition of No FOK items. Finally, gamma correlations are analysed, and compared to chance performance. This structure of analysis is repeated for all FOK experiments in Chapters 2 and 3.

To examine possible bias in responding, the proportion of unrecalled items that were assigned a Yes FOK response was calculated (see Table 2.2 for descriptives). A 2 (age) x 2 (task) repeated measures ANOVA revealed no effect of age, $F(1,50) = 1.456$, $p = .233$, $\eta^2 = .028$. A main effect of task was observed, $F(1,50) = 16.814$, $p < .001$, $\eta^2 = .252$, with a greater proportion of Yes FOKs in the Episodic task than the Semantic task. No interaction was present, $F(1,50) = 2.489$, $p = .121$, $\eta^2 = .047$.

Examination of subsequent recognition for items assigned Yes FOKs and No FOKs can provide an indication as to the accuracy of the FOK judgement. It would be expected that a larger proportion of items given a Yes FOK would go on to be recognised than items given a No FOK. The number of correctly recognised items assigned a Yes FOK was divided by the total number of Yes FOK responses to give proportion accuracy of Yes FOKs. Likewise, the number of correctly recognised items given a No FOK was divided by the total number of No FOK responses to give proportion accuracy of No FOKs. The recognition accuracy of Yes FOK responses and No FOK responses was submitted to a 2 (age) x 2 (task) x 2 (status: Yes FOK and

Table 2.2: Mean proportions of unrecalled items assigned a Yes FOK response for Semantic and Episodic FOKs in young and older adults

		Young adults	Older adults
		M (SD)	***M (SD)***
Liberal Criterion	Semantic	0.301 (0.158)	0.446 (0.141)
	Episodic	0.561 (0.281)	0.561 (0.411)
Strict Criterion	Semantic	0.301 (0.158)	0.442 (0.146)
	Episodic	0.572 (0.277)	0.545 (0.432)

Note: *N=35* for young adults in both analyses. For older adults, *N=17* in the liberal criterion and *N=16* in the strict criterion.

No FOK) repeated measures ANOVA (see Figure 2.2). The key analysis here is whether the three-way interaction between these variables reaches significance. If it does, this could indicate the expected deficit in older adults' Episodic FOKs. A main effect of age was present, $F(1,35) = 7.885$, $p = .008$, $\eta^2 = .184$, consistent with the foregoing analysis. No main effect of task was present, $F(1,35) = 0.963$, $p = .333$, $\eta^2 = .027$. A main effect of status was observed, $F(1,35) = 20.484$, $p < .001$, $\eta^2 = .369$, with greater recognition accuracy for Yes FOKs than No FOKs. This indicates diagnostic accuracy in participants' FOK predictions. No interactions were present (age and task: F<1; age and status: F<1; task and status: F<1; age, task and status: F<1). At this stage, the older adults actually have superior memory performance. In turn, both groups' performance is higher for items assigned to the Yes FOK category. Results thus far are indicative of FOK accuracy for both Episodic and Semantic FOK in both age groups.

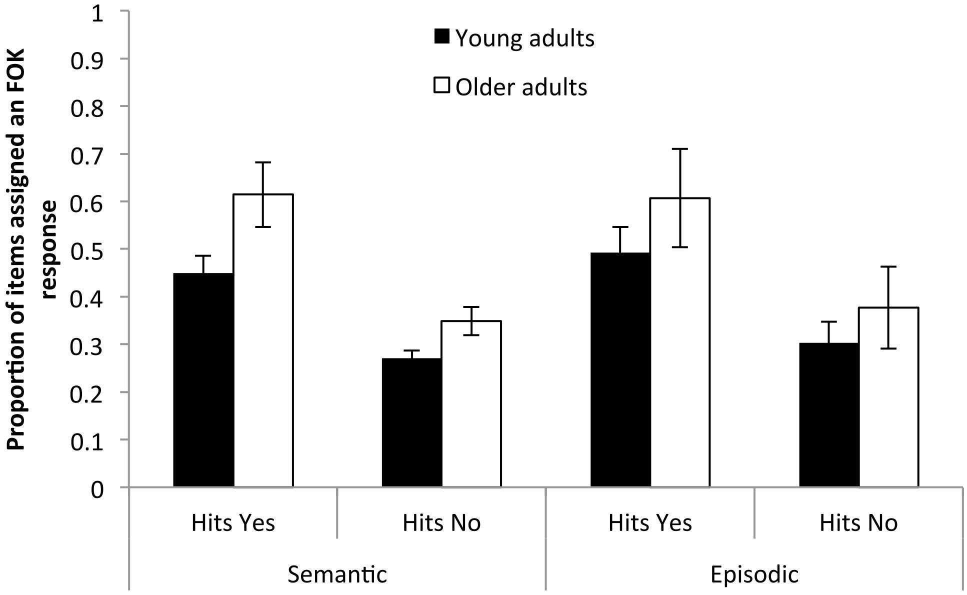


Figure 2.2: Correct recognition of Semantic and Episodic items assigned Yes and No FOKs for young and older adults in the liberal criterion

There has been considerable debate as to the most appropriate measure of FOK accuracy (Nelson, 1984; Schraw, 1995). The association between the Yes/No FOK judgement and recognition performance as obtained in the present study is most appropriately measured with the Goodman-Kruskal gamma correlation (Nelson, 1984; Schraw, 1995; Wright, 1996; see Masson and Rotello, 2009, for discussion of measures when using rating scales for FOK). Four possible outcomes can occur with a binary FOK: (a) correct recognition for Yes FOKs, (b) incorrect recognition for Yes FOKs, (c) correct recognition for No FOKs, and (d) incorrect recognition for No FOKs. The gamma correlation compares the proportion of correct predictions (a and d) to incorrect predictions (b and c). This leads to a coefficient ranging from 1 to -1, with large positive values indicating a close relationship between FOK judgement and recognition, large negative values indicating an inverse relationship between FOKs and recognition, and values close to zero showing chance performance or guessing. However, gamma is undefined when two of the four possible outcomes (a, b, c, d) are equal to 0. Therefore, an adjusted gamma score was calculated following Snodgrass and Corwin's (1988) recommendations (see Souchay et al. 2000, 2007), whereby 0.5 was added to each frequency and then divided by N+1, where N is the number of judgements. This strategy of analysis was followed for all experiments in Chapter 2 and Chapter 3.

Mean gamma correlations for each age group are shown in Table 2.3. A 2 (age) x 2 (task) repeated measures ANOVA was used to examine FOK accuracy as measured by gamma. No effect of age was found, $F<1$, however a marginal effect of task was observed, $F(1,50) = 3.360$, $p = .073$, $\eta^2 = .063$; FOK accuracy was higher for the Semantic task than the Episodic task. No interaction was present, $F(1,50) = 1.954$, $p = .168$, $\eta^2 = .038$. Gamma correlations were also tested using a one tailed t-test against zero to establish if accuracy was greater than chance. In the young adult group, gamma correlations were significantly above chance for both the Semantic, $t(34) = 6.743$, $p < .001$, and Episodic, $t(34) = 4.295$, $p < .001$, task. For the older adults, performance was significantly above chance for the Semantic task, $t(20) = 4.904$, $p < .001$, and marginally above chance for the Episodic task, $t(16) = 2.108$, $p = .051$.

Pearson's correlations between gamma and recall accuracy were examined to further explore the relationship between metacognitive accuracy and memory performance. For young adults, no relationship was observed between gamma and recall for the Semantic task, $r(35) = .031$, $p = .858$, or for the Episodic task, $r(35) = -.109$, $p = .534$. Likewise, for older adults, no correlations were present between gamma and recall for either the Semantic, $r(17) = .403$, $p = .109$, or Episodic, $r(17) = .284$, $p = .269$, tasks.

In sum, analysis of the gamma correlations does not appear to show an age effect on FOK accuracy for either the Semantic or Episodic tasks. Both young and older adults have higher resolution in the Semantic task, and all gamma correlations are above chance accuracy, although this is marginal in the Episodic task for older adults. In addition, no relationship is observed between recall and FOK accuracy for either task in either age group.

2.2.3.2 Strict criterion

Five older adult participants were removed from the analysis due to no errors of omission occurring. Therefore no metacognitive data could be analysed for these participants. The subsequent analysis involved 35 young adults (age M = 20.23, SD = 2.89) and 16 older adults (age M = 68.75, SD = 8.20). The following analysis is a direct repeat of the structure and rationale for the liberal criterion. As previously, proportions for the Semantic trial are calculated by dividing responses by 60. Proportions for the Episodic trial are dependent on the performance of each individual participant at the Semantic recognition task. An average of 33.137 (SD = 8.412) items comprised the Episodic task. Proportions were calculated as previously described for the liberal criterion.

Memory

As can be seen in Figure 2.3, it would appear that older adults show a distinct advantage in Semantic memory while Episodic memory is more similar between the two age groups. A main effect of age was found on recall performance, $F(1,49) = 18.403$, $p < .001$, $\eta^2 = .273$, again with older adults showing greater accuracy than young adults. Similarly, recall on the Episodic task was also significantly greater than recall for the Semantic task, $F(1,49) = 347.532$, $p = .001$, $\eta^2 = .876$. An interaction between age and task, $F(1,49) = 7.339$, $p = .009$, $\eta^2 = .130$, showed that although both groups were significantly more accurate in the Episodic task than the Semantic task (young adults: $t(34) = 22.579$, $p < .001$; older adults: $t(15) = 7.483$, $p < .001$), at the Semantic trial older adults are at a greater level of accuracy than young adults, $t(49) = 6.501$, $p < .001$, whereas at the Episodic trial no age difference is present $t(49) = 1.581$, $p = .120$.

Recognition performance also showed a main effect of age, $F(1,49) = 8.056$, $p = .007$, $\eta^2 = .141$, with older adults recognising a greater proportion of items correctly than young adults as previously. A main effect of task was also present, $F(1,49) = 118.332$, $p < .001$, $\eta^2 = .707$, again with Episodic task recognition being higher than that of Semantic task recognition. An interaction between age and task mirrored that of the interaction in recall performance, $F(1,49) = 31.370$, $p < .001$, $\eta^2 = .390$, with greater improvement between tasks in the young adult group, $t(34) = 16.803$, $p < .001$, than in the older adult group, $t(15) = 2.575$, $p = .021$.

As previously, errors of omission and commission were analysed (see Table 2.1). Errors of omission revealed a similar pattern. A main effect of age was due to older adults giving fewer errors of omission than young adults, $F(1,49) = 37.081$, $p < .001$, $\eta^2 = .431$, and a main effect of task was present with more omissions during the Semantic task than the Episodic, $F(1,49) = 383.717$, $p < .001$, $\eta^2 = .887$. An interaction between age and task, $F(1,49) = 7.020$, $p = .011$, $\eta^2 = .125$ again appeared to be driven by greater disparity at the Semantic trial than the Episodic trial (follow up independent t-tests: Semantic task $t(49) = 6.020$, $p < .001$, Episodic task $t(49) = 3.559$, $p < .001$; follow up paired samples t-tests: young adults $t(34) = 23.569$, $p < .001$, older adults $t(15) = 7.938$, $p < .001$). Errors of commission also show a similar pattern to that observed under the liberal criterion, with a main effect of age, $F(1, 49) = 21.128$, $p < .001$, $\eta^2 = .301$, and no effect of task, $F<1$, or interaction between age and task, $F<1$.

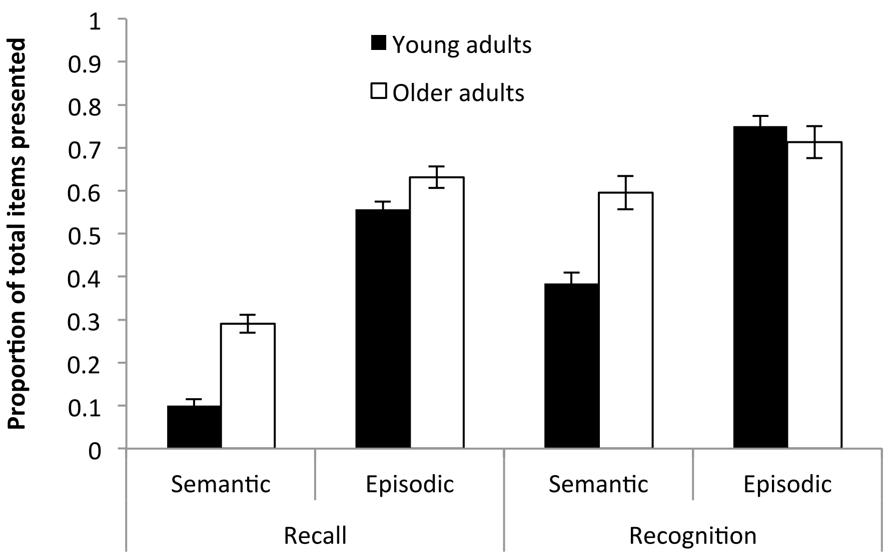


Figure 2.3: Proportion of total items presented which were correctly recalled and recognised by young and older adults in each condition based on the strict criterion

Metacognition

A 2 (age) x 2 (task) repeated measures ANOVA on the proportion of Yes FOKs given revealed no effect of age, $F(1,49) = 0.889$, $p = .350$, $\eta^2 = .018$ (see Table 2.2). A main effect of task was present, $F(1,49) = 14.646$, $p < .001$, $\eta^2 = .230$, with more Yes FOKs given in the Episodic task than in the Semantic task, as per the results with the liberal criterion. A marginal interaction between age and task, $F(1,49) = 2.954$, $p = .092$, $\eta^2 = .057$, was found due to young adults increasing Yes responding between tasks, $t(34) = 5.837$, $p < .001$, whereas older adults do not, $t(15) = 0.995$, $p = .336$. In addition, older adults gave more Yes FOK responses at the Semantic task than young adults, $t(49) = 3.047$, $p = .004$, whereas no age differences in responding were present for the Episodic task, $t(49) = 0.224$, $p = .825$.

Accuracy of Yes and No FOKs was analysed with a 2 (age) x 2 (task) x 2 (FOK status) repeated measures ANOVA. No effect of age, $F<1$, or task, $F(1,33) = 1.078$, $p = .307$, $\eta^2 = .032$, was present. A main effect of status, $F(1,33) = 48.987$, $p < .001$, $\eta^2 = .597$, confirms that the FOK judgements do show some diagnostic accuracy, with Yes FOKs resulting in greater recognition accuracy than No FOKs (see Figure 2.4). An interaction between age and task, $F(1,33) = 7.912$, $p = .008$, $\eta^2 = .193$, indicates that older adults show a slight decrease in performance from the Semantic task to the Episodic task. In contrast, young adults show an increase in performance, with Semantic task performance lower than that of Episodic task performance.

A marginal interaction was also found between age and FOK status, $F(1,33) = 3.317$, $p = .078$, $\eta^2 = .091$. Older adults actually show a greater disparity between Yes and No FOK performance than young adults, indicating greater metacognitive accuracy. No other interactions were significant (task and FOK status: $F(1,33) = 2.006$, $p = .166$, $\eta^2 = .057$; age, task and FOK status: $F<1$).

Table 2.3: Gamma correlations between FOK and recognition for Semantic and Episodic memory in young and older adults.

		Young adults	Older adults
		M (SD)	*M (SD)*
Liberal criterion	Semantic Gamma	0.345 (0.303)	0.466 (0.213)
	Episodic Gamma	0.313 (0.431)	0.223 (0.436)
Strict criterion	Semantic Gamma	0.345 (0.303)	0.455 (0.214)
	Episodic Gamma	0.477 (0.408)	0.177 (0.459)

Note: *N=39* for young adults in both analyses. For older adults, *N=17* in the liberal criterion and *N=16* in the strict criterion.

A 2 (age) x 2 (task) repeated measures ANOVA on gamma correlations revealed no effect of age, $F(1, 49) = 1.560$, $p = .218$, $\eta^2 = .031$, or of task, $F<1$, on FOK accuracy (see Table 2.3). A significant interaction was found however, $F(1,49) = 7.044$, $p = .011$, $\eta^2 = .126$. Young adults show no difference in accuracy between the Semantic and Episodic tasks, $t(34) = 1.490$, $p = .145$, whereas older adults show significantly lower accuracy in the Episodic task compared to the Semantic task, $t(15) = 2.284$, $p = .037$. In addition, while no effect of age was found for the Semantic task, $t(49) = 1.299$, $p = .200$, for the Episodic task older adults were significantly lower in accuracy than the young adults, $t(49) = 2.341$, $p = .023$. Gamma correlations were also compared to chance performance to establish if predictions were indicative of recognition accuracy. For young adults, both Semantic, $t(34) = 6.743$, $p < .001$, and Episodic, $t(34) = 6.915$, $p < .001$, gammas were greater than chance. However, for older adults, only Semantic gamma was above chance, $t(20) = 4.904$, $p < .001$; Episodic gamma, $t(15) = 1.542$, $p = .144$.

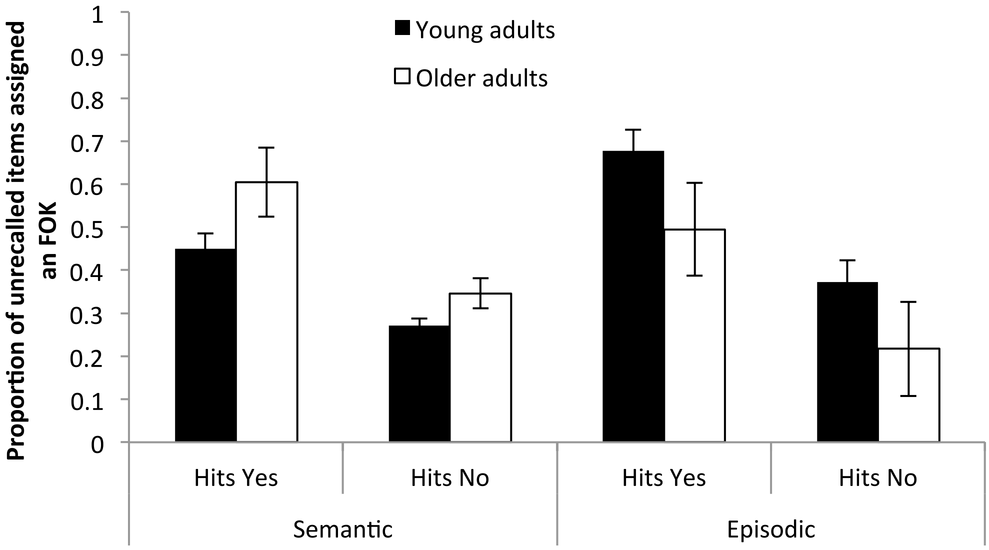


Figure 2.4: Correct recognition of Semantic and Episodic items assigned Yes and No FOKs for young and older adults in the strict criterion.

Pearson's correlations again revealed no relationship between recall accuracy and gamma correlations for young adults in the Semantic, $r(35) = .031$, $p = .858$, or Episodic, $r(35) = -.266$, $p = .122$, tasks. Similarly, no correlations were observed for Semantic, $r(16) = .386$, $p = .140$, or Episodic, $r(16) = .014$, $p = .959$, tasks within the older adult group.

To summarise, the results from the gamma correlation analysis indicate a selective impairment of Episodic FOK accuracy in older adults. This occurs independently of any age effects on Episodic memory accuracy. The findings of the liberal criterion and the strict criterion do, for the most part, agree. No age-related deficit is found for measures of memory accuracy, nor for reliance on Yes or No FOK responding. Likewise, for both young and older adults Yes FOKs are more likely to be associated with correct recognition than No FOKs. The two analyses do diverge slightly when it comes to gamma. For the liberal criterion, no age effect is found, although older adults are only just above chance in the Episodic task. For the strict criterion in contrast, a clear age effect is shown, with older adults having preserved Semantic FOK accuracy but impaired Episodic FOK accuracy.

2.2.3.3 Recollection

Only items which were correctly recognised were included in the recollection analysis, therefore the number of items used varied by participant in accordance with their recognition accuracy. Proportions were subsequently calculated to allow comparison between individuals. The number of items assigned an R response was divided by the total number of correctly recognised items to give proportion R responses. Proportion K responses were also calculated this way. The proportion of R and K responses for hits was analysed using a 2 (age) x 2 (state: R and K) repeated measures ANOVA (see Figure 2.5). No effect of age was found, $F(1,54) = 0.077$, $p = .783$, $\eta^2 = .001$. A main effect of state was shown, $F(1,54) = 67.741$, $p < .001$, $\eta^2 = .556$, with more R responses given then K responses. An interaction between age and state, $F(1,54) = 27.043$, $p < .001$, $\eta^2 = .334$, was also shown. Follow up t-tests revealed that, for young adults, significantly more R responses were given than K responses, $t(34) = 11.200$, $p <.001$, whereas for older adults no difference could be detected between R and K responding, $t(20) = 1.853$, $p = .079$. Furthermore, young adults give significantly more R responses than older adults, $t(54) = 3.909$, $p < .001$, whereas older adults give significantly more K responses than young adults, $t(54) = 5.512$, $p < .001$. Thus older adults show lower levels of R responding, as expected.

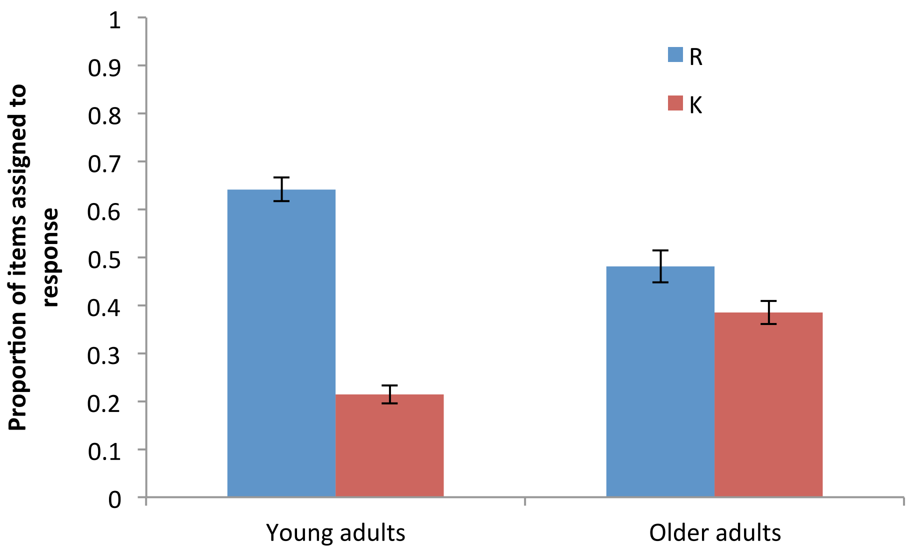


Figure 2.5: Proportion of R and K responses given for correctly recognised items only by young and older adults

In addition to examining all responses together, items can also be split into Semantic and Episodic as above for the memory and metacognitive analyses. The expectation here would be that items classed as Semantic would be associated with increased Know responses whereas items classed as Episodic would be associated with increased Remember responses. Both liberal and strict criterion-based analyses were conducted using a 2 (age) x 2 (task: Semantic and Episodic) x 2 (state: R and K) repeated measures ANOVA. The liberal criterion analysis (see Table 2.4) revealed no effect of age, $F(1,49) = 1.097$, $p = .300$, $\eta^2 = .022$, but a main effect of task, $F(1,49) = 142.159$, $p < .001$, $\eta^2 = .744$, with the Semantic task leading to a higher rate of reporting than the Episodic task. A main effect of state was also found, $F(1,49) = 16.866$, $p < .001$, $\eta^2 = .256$, with a greater proportion of K responses than R responses. An interaction between task and status was also shown, $F(1,49) = 248.512$, $p < .001$, $\eta^2 = .835$, with the Semantic task showing greater K responses than R responses while the Episodic task showed a greater proportion of R responses than K responses. No other interactions reached significance (age and task: $F<1$; age and state: $F<1$; age, task and state: $F<1$).

Table 2.4: Means (*SD*) of proportion of R and K responses according to Semantic-Episodic distinction. Liberal criterion.

	Semantic		**Episodic**	
	R	**K**	**R**	**K**
Young adults	0.136 (0.252)	0.856 (0.263)	0.598 (0.161)	0.122 (0.083)
Older adults	0.064 (0.135)	0.884 (0.185)	0.58 (0.219)	0.12 (0.088)

For the strict criterion (see Table 2.5), again no effect of age is shown, $F<1$, nor is an effect of task present, $F(1,49) = 2.279$, $p = .138$, $\eta^2 = .044$. A main effect of state is repeated, $F(1,49) = 94.459$, $p < .001$, $\eta^2 = .658$, with a greater proportion of R responses than K responses. No significant interactions were detected between age and task, $F<1$, or between age and state, $F(1,49) = 2.061$, $p = .157$, $\eta^2 = .040$. However, an interaction was found between task and state, $F(1,49) = 178.852$, $p <$

.001, $\eta^2 = .785$, with the Semantic task showing similar levels of R and K responding while the Episodic task showed much greater R responding than K responding. A three way interaction between age, task and state was also present, $F(1,49) = 4.846$, $p = .032$, $\eta^2 = .090$. Both age groups show the same pattern of responding in the Episodic task, with higher levels of R responses than K responses. For the Semantic task, young adults show quite similar levels of R and K responding, whereas older adults show greater K responding than R.

Table 2.5: Means (*SD*) of proportion of R and K responses according to Semantic-Episodic distinction. Strict criterion.

	Semantic		Episodic	
	R	**K**	**R**	**K**
Young adults	0.438 (0.210)	0.443 (0.203)	0.787 (0.174)	0.048 (0.078)
Older adults	0.304 (0.111)	0.549 (0.146)	0.813 (0.276)	0.022 (0.072)

The relationship between RK and FOK can be assessed by calculating a gamma correlation between the two judgements (Hicks and Marsh, 2002). However for the present data set, insufficient items are available for this correlation to be examined for each individual. Unlike the FOK and recognition gamma described above, no correction is available to allow estimation of the correlation when there are missing data points. To give some indication as to the relationship of the two measures within each age group, a single gamma correlation measure was calculated based on total responses in each category within each age group. For both young and older adults, gamma correlations indicated a strong relationship between the two subjective experiences, with young adults obtaining an overall correlation of 0.809 and older adults obtaining a correlation of 0.809 also.

Due to the difficulty of calculating an RK-FOK gamma, standard Pearson's correlations were examined. The proportion of items assigned R and K responses were correlated against Semantic FOK accuracy and Episodic FOK accuracy based on both the liberal and strict criteria for young and older adults separately (see Table 2.6). No correlations were observed between the two subjective experiences for young or older adults. This analysis therefore suggests that the two processes are not related.

Table 2.6: Correlation matrix to examine the relationship between Remember and Know responding and FOK accuracy in young and older adults

			Semantic gamma	Episodic gamma (liberal)	Episodic gamma (strict)
Young adults	**Remember**	***r***	-.217	.122	.228
		p	.211	.486	.188
	Know	***r***	.304	-.044	-.168
		p	.076	.800	.335
Older adults	**Remember**	***r***	-.117	.267	-.114
		p	.655	.301	.674
	Know	***r***	.306	.077	.403
		p	.233	.768	.122

2.2.4 Discussion

The aim of Experiment 2.1 was to examine the effect of ageing on semantic and episodic FOK judgements within a single task. Due to the design of the task, the items included within the Episodic measure could be identified with either a liberal criterion (i.e. all items not correctly recalled during the Semantic task) or with a strict criterion (i.e. all items not correctly recognised during the Semantic task). With regards to the liberal criterion, both measures of memory performance show a similar pattern: older adults outperform young adults generally, and Episodic accuracy is greater than that of Semantic. The greater accuracy in older adults is not, perhaps, surprising if it is driven mainly by the Semantic performance, with previous findings suggesting older adults may outperform young adults on certain tests of semantic knowledge (e.g. Verhaeghen, 2003). Indeed, even within previous assessments of semantic memory in the context of the FOK paradigm, evidence has been found for superior accuracy in older adults (Butterfield et al., 1988; Marquié, Jourdan-Boddaert, & Huet, 2002), though this has typically been confined to recognition performance only. However, closer consideration of the interaction between age and task indicates that Semantic memory performance is not the sole contributor to the main effect of age for recall performance. Although both groups improve in accuracy between Semantic and Episodic measures, young adults remain below the level of accuracy that the older adults are able to achieve. Thus, for this task, we have the unusual finding that older adults are able to accurately recall a higher proportion of episodic information than young adults, a point that we will return to later.

The metacognitive data suggest that young and older adults utilise the FOK judgement in a similar way, with no differences in the tendency to give positive FOK responses. This could represent similar levels of confidence in future performance by the two groups, and greater confidence in recognising unrecalled items for the Episodic task than the Semantic task. Analysis of the hit rates for Yes and No FOKs indicates predictive accuracy in judgements for both groups, as a greater proportion of items given Yes FOKs are recognised than items given No FOKs.

The gamma correlation provides an unbiased measure of accuracy (see Chapter 4 for a more detailed discussion of measures of FOK accuracy). To establish whether there is a selective impairment of episodic FOK accuracy in ageing, the key analysis to consider is whether an interaction is observed in the ANOVA for gamma correlations. If, as suggested by Souchay et al. (2007) and others, older adults are still able to make semantic FOK judgements but are less able to make episodic FOKs, this should be identified through an interaction of task type and age. When considering the liberal criterion to define which items are classed as Episodic, this interaction fails to reach significance. There is some indication though that older adults do find it more difficult to make predictive judgements for the Episodic task, as shown with the comparisons of gamma against chance performance. Young adults are reliably above chance for both Semantic and Episodic tasks, whereas the older adults are above chance for the Semantic task and only marginally above chance for the Episodic task. Although not conclusive, this may imply a selective deficit in episodic predictive accuracy for the older adult group.

One potential problem with the analysis presented above is the definition of which items are episodic. The liberal criterion is based on correct recall of items. If a question is correctly recalled during the Semantic task, then it is removed from the Episodic analysis. However, it is possible that participants do know the correct response during the Semantic task, but they are unable to access the answer during a recall test. There could therefore be contamination of Semantic items within the Episodic measure for the liberal criterion. A more conservative criterion would further limit the influence of Semantic memory, giving a more 'pure' estimate of Episodic processes. The strict criterion was thus based on recognition performance during the Semantic task. Presumably items to which the answer is known but not retrievable for recall would be more likely to be accessed via a recognition test, as evidenced in numerous studies where an advantage is found for recognition memory over recall (e.g. Mandler, 1967). We can therefore state with greater confidence that the remaining items are episodic in nature.

With regards to memory performance, results based on the strict criterion largely mirror those of the liberal: older adults outperform young adults, and Episodic memory is better than that of Semantic. Again, young adults show a greater difference between the Semantic and the Episodic tasks, but in this instance no age effect remains at the Episodic test. Young adults and older adults are at similar levels of accuracy. With regards to the FOK, no bias in responding was detected, and Yes FOKs were associated with more hits than No FOKs, suggesting some level of predictive accuracy for both age groups. The critical analysis is again the presence or absence of an interaction between age and task for the gamma correlations. For the strict criterion, this interaction does reach significance, and in the direction proposed by Souchay et al. (2007) and others. In the Semantic task, no age effect is observed in gamma levels, whereas for the Episodic task older adults are significantly less accurate than young adults. In fact, gamma correlations are no different from chance for older adults in the Episodic task only. This consequently demonstrates a selective impairment in episodic FOK accuracy in older adults.

The data presented lend further weight to previous findings of episodic FOK deficits in ageing (Perrotin et al., 2006; Souchay et al., 2000, 2007; Thomas et al., 2011). The presence of a deficit limited to episodic tasks only can be inferred from a variety of closely related research areas. Age-related changes in the structure of the brain in particular have been shown to influence episodic FOKs, either directly or via the influence on related cognitive processes. The frontal lobes are especially vulnerable to age-related atrophy (Dennis et al., 2008; Raz et al., 2005). Neuroimaging studies have identified the greater importance of the frontal lobes in making episodic FOK judgements as opposed to semantic FOKs (Kikyo & Miyashita, 2004; Kikyo, Ohki, & Miyashita, 2002; Anat Maril et al., 2003, 2005; Reggev et al., 2011; Schnyer et al., 2005), so damage to this area of the brain, as occurs in normal ageing, will disproportionately affect episodic FOKs compared to semantic FOKs. Similarly, changes to frontal lobe function related to ageing have been found to impact executive function processes (Goh et al., 2013; Treitz et al., 2007), which in turn has been found to be a major contributor to episodic FOK accuracy (Perrotin et al., 2006; Souchay et al., 2004, 2000; Souchay & Isingrini, 2004). Impairments in executive function due to frontal degeneration associated with ageing therefore lead to deficits in episodic FOK. A final line of evidence from which a selective episodic FOK deficit can be inferred is that of recollection. Older adults report less recollective experience (Bugaiska et al., 2007; Bunce, 2003; Perfect & Dasgupta, 1997), and lower levels of remembering are associated with lower accuracy of FOK (Hicks & Marsh, 2002). The relationship between recollection and episodic FOK accuracy has been further demonstrated specifically within ageing, with both subjective (Remember-Know: Souchay et al., 2007) and objective (source memory:

Thomas et al., 2011) measures of recollection. As recollection deteriorates, so too does episodic FOK accuracy. Taken together, these findings suggest that ageing would affect episodic FOK accuracy in isolation, leaving semantic FOK preserved, findings which are further supported by the present study.

As well as the inferential hypothesis described above, a second hypothesis has been suggested to account for the episodic FOK data. The Memory Constraint Hypothesis (MCH: Hertzog et al. 2010) states that the accuracy of episodic FOKs is dependent on the quality of the original encoding. If a deficit occurs at encoding there will be an insufficient quantity and quality of partial information available at the retrieval attempt to support a suitable FOK judgement, leading to lower accuracy. It is true that some findings of episodic FOK impairment in ageing also show episodic memory impairments within the older adult participants (Perrotin et al., 2006; Souchay et al., 2000, 2007); however, not all of them do (Thomas et al., 2011). Indeed, the present results cannot be suitably explained by the MCH, whether considering the liberal or strict criterion to define Episodic memory. For the liberal criterion, superior Episodic memory was observed in the older adult group, and yet no related benefit was shown on Episodic FOK accuracy. Likewise in the strict criterion, older adults were at a similar level of Episodic memory accuracy to young adults. Yet when considering FOK accuracy, older adults were found to be at a significantly lower level than young adults. For neither of these criteria did the memory and FOK results mirror each other, as the MCH would predict. In addition, no relationship was observed between memory performance and FOK accuracy for either age group in either memory task. This therefore questions the utility of the MCH in explaining the disparate findings within the ageing and FOK literature.

With regards to recollection, no age effect was observed in the present study. When examining all items, the presence of the interaction may be considered an indication of an age-related deficit in recollection, due to older adults showing similar levels of R and K responding whereas young adults discriminate R and K responses to a greater extent. However, this interaction is not borne out in follow up analyses where the Semantic and Episodic tasks are considered separately. The relationship between recollection and FOK accuracy within the current task could not unfortunately be assessed with gamma due to missing data points (see Chapter 4). A gamma correlation was calculated based on the sum responses for each age group, indicating a large positive relationship between RK responding and FOK judgements for all participants. However Pearson's correlations did not indicate any relationship between the two subjective experiences in young or older adults.

The advantage of the paradigm used in Experiment 2.1 is that FOK and RK responses are obtained for the same materials unlike in Souchay et al. (2007), allowing greater control of task and item effects on these judgements. This therefore reduces the noise in the data and provides a more accurate consideration of the relationship. The lack of relationship found between FOK and RK at the individual level may suggest that the two experiences are independent and do not share similar bases or processes. However, the reliance on Pearson's correlations rather than gamma, although necessary in this case, is not ideal and could be responsible for the absence of correlation. The relative nature of the gamma is more suitable to quantify the relationship between recollection and FOK. Thus the current findings should be interpreted with caution.

The recollection measure further allowed confirmation that the Semantic and Episodic tasks were approximating true measures of semantic and episodic memory. Tulving (1972) proposed that episodic memory is characterised by the subjective experience of remembering, the retrieval of contextual details in addition to the target memory leading to a richer memory experience. In contrast, semantic memory is characterised by the subjective experience of knowing, the ability to retrieve the answer. These experiences can then be estimated through the RK paradigm, with higher R responses associated with episodic memory and higher K responses associated with semantic memory. This pattern of responding is shown both for the liberal and strict criterion used to define Episodic memory in the current task, adding support to the assumptions used to distinguish Semantic and Episodic processes.

The present study has successfully evaluated the effect of ageing on semantic and episodic FOK within the same participants and controlling for task characteristics. It has demonstrated a selective age deficit in episodic FOK accuracy, with semantic FOK accuracy preserved for older adults. However, the distinction of what is classed as Episodic does limit interpretation of the results to some extent. The liberal criterion of removing only those items which are correctly recalled is perhaps too liberal, as some items which are not recalled may nonetheless be known by the participants. The strict criterion would therefore be more appropriate as an approximation of Episodic memory, limiting the influence of Semantic knowledge to the greatest extent possible within a repeated measures design as this. However, with the strict criterion there is a problem of reduced data points for analysis. The high levels of Semantic recognition in the older adult group are of particular concern, as this reduces the items available for calculating Episodic FOK and its accuracy to unsuitably low levels. The general knowledge questions used in the present study do not offer enough of a challenge to the older adult sample. A replication of the paradigm is needed, with an increased level of difficulty to ensure sufficient data points when using a strict criterion definition of Episodic. This would allow greater confidence in the analysis and interpretation of the effect of age on semantic and episodic FOK processes.

In conclusion, the present study has shown that an age effect is present in FOK accuracy. Although semantic FOKs are preserved in healthy ageing, episodic FOKs are impaired, leading to lower diagnostic accuracy in predictions of future recognition. The present paradigm provides a promising new avenue of examining these two processes with greater control of sources of variance in metacognitive judgements, such as individual biases in responding and the influences of task characteristics.

2.3 Experiment 2.2: Language learning with repeated learning trials

2.3.1 Introduction

In addition to the more widely used general knowledge tasks and word pair learning tasks described previously, FOK has also been assessed with language learning paradigms. This single paradigm can be applied to either a semantic FOK task by using a second language that the participant has some level of knowledge of (Peynircioğlu & Tekcan, 2000) or as an episodic task by using a second language that the participant has not encountered previously (Nelson & Dunlosky, 1994). The aim of Experiment 2.2 was to use a language learning paradigm in a combined semantic and episodic task. Participants were presented with a word in the second language and asked for the English translation. FOK judgements were then made, and a recognition test assessed the accuracy of the FOK predictions, thus measuring semantic FOK. Correct translations were then provided for learning, and a second translation, FOK and recognition procedure was completed whereby items not correctly translated in the semantic task were used to measure episodic FOK accuracy. In both cases, the prompt to elicit the recall and FOK judgements was identical i.e. what is the English translation? Thus, in addition to measuring semantic and episodic FOK accuracy within the same participants, it was also possible to do this with the same target items and with the same cue prompts, further limiting the influence of task characteristics as was achieved in Experiment 2.1.

The issue of expertise is of concern when utilising this type of task due to the widespread finding that memory performance is enhanced with expertise (e.g. Engle & Bukstel, 1978; Vicente & Wang, 1998). The relationship between memory and FOK is complex, and thus it is important to consider the influence of prior knowledge on FOK predictions. To date, only three studies have directly examined the role of expertise on FOK. Roberts & Rhodes (1989) observed no difference between FOK accuracy of experts and non-experts for a general knowledge quiz. However, in this instance expertise was ascertained by using self-report: participants were asked to rate their own ability at answering general knowledge questions. Although self-rated experts did indeed recall more items than non-experts, the actual level of ability for each group in unknown. Marquié & Huet (2000) also found no effect of self-efficacy beliefs on FOK accuracy in young and older adults for computer knowledge, though again the level of expertise was ascertained by self-report. Peynircioğlu & Tekcan (2000) assessed the impact of language proficiency on FOK judgements for Turkish-English translations. Participants were native Turkish speakers who were required to pass a proficient examination in English in order to attend a university course, thereby allowing an independent assessment of expertise. Results replicated those of the Roberts & Rhodes (1989) study. Participants with higher expertise did correctly translate more items than participants with lower expertise, however no difference was found in FOK accuracy between the two groups. Based on these three studies, it would appear that expertise does not impact the accuracy of FOK judgements despite its well documented effects on memory performance.

The representation of language in bilingual memory is also an important consideration when using a translation paradigm. According to the revised hierarchical model (Kroll & Stewart, 1994) words for the first (or native) language (L1) are stored in a separate lexical memory system to words in the second language (L2). The semantic features of the words are then stored at a further, conceptual level. During a translation task, highly proficient bilinguals are able to access conceptual details of L2 words directly. In contrast, novice bilinguals have to rely on word-to-word associations between L2 and L1 to indirectly access conceptual details. This reflects a shift from episodic-based access to meaning in novice bilinguals, to semantic-based access to meaning in fluent bilinguals. In order to maximise the use of episodic memory processes, it is therefore essential that translation occurs from the L2 to L1 direction, relying on word-to-word associations and removing the influence of semantic links between the L1 word and the concept. To further improve the episodic aspects of the task, no conceptual or semantic details will be provided at encoding. With only the L1 and equivalent L2 words provided, participants will be encouraged to utilise word-to-word associations only to learn the correct translations (Kroll & Curley, 1988).

The main motivation for using a language learning paradigm is that it allowed a further test of the MCH. If sufficient words are used in the translation task, there is the opportunity to administer repeated learning trials and obtain repeated FOK judgements for those trials. The greater opportunity to encode information will also increase the quality of the encoding, which the MCH predicts will also increase FOK accuracy. A number of previous studies have examined the effect of increased encoding on FOK judgements, either by altering the learning conditions or by requiring participants to reach a criterion level of recall before measuring memory and FOK after a four week delay. Hart (1967) manipulated the number of presentations of cue-target pairs, finding no effect of one, two or three repetitions on FOK accuracy despite increased memory performance with increased repetitions. Likewise, Schacter (1983) presented items for either 1 ½ or 5 seconds, finding no impact of length of presentation on FOK accuracy, despite increased memory performance with increased time. However, neither author used Gamma to measure accuracy of FOKs in these tasks. Instead, the proportion of Hits for Yes FOKs was analysed, a measure which is prone to bias from increased memory performance (for a more detailed discussion on the advantages and disadvantages of various measure of metacognitive accuracy, see Chapter 4).

In contrast, tasks using learning to criterion have used the Gamma correlation to assess accuracy. Nelson, Leonesio, Shimamura, Landwehr, & Narens (1982) required participants to correctly recall items either one, two, or four times. Memory and FOK were then assessed four weeks later, with the observation that FOK accuracy was significantly greater for those items correctly recalled four times previously than two or one times. Carroll & Simington (1986) attempted to replicate this result with a different criterion of one or three correct recalls, however no effect on FOK accuracy was found.

A key difference between these two studies was the design used. Nelson et al. (1982) used a within subjects design, whereas Carroll and Simington (1986) used a between subjects design. Carroll & Nelson (1993) proposed that the between subjects design allowed for too great an effect of individual differences on the FOK, thereby removing the effect of the encoding manipulation. They emphasised the difference between encoding conditions by requiring either one correct recall or six correct recalls, and compared both a between and within subjects design. They observed no effects of FOK accuracy for the between subjects design, but reliable effects of increased learning on FOK accuracy for the within subjects design. By using the same participants in each encoding condition, the variability in decision criteria for Yes and No FOKs is reduced, thereby allowing assessment of the effect of the encoding manipulation. With between subjects designs, subjective ratings are not sensitive to the effects of the manipulation due to increased variance rather than due to no effect occurring. This artefact of the design of the study is also an issue with the studies by Hart (1967) and Schacter (1983), both of whom used a between subjects design for their tasks. More recently, Hertzog et al. (2010) examined the effect of repeated exposures to cue-target pairs in young and older adults in a within subjects comparison. Gamma correlations showed a reliable effect of increased learning in both young and older adults: as the number of exposures increased, so too did memory accuracy, and FOK accuracy.

The effect of increased learning would therefore appear to not only improve memory but also improve FOK accuracy. This perfectly demonstrates the MCH, that increased encoding allows access to more diagnostic information on which to base FOK judgements, thus leading to greater FOK accuracy. To date, only Hertzog et al. (2010) appear to have shown this effect on FOK with repeated learning trials using an unbiased accuracy measure. The present study therefore aims to replicate this manipulation of encoding to further establish the utility of the MCH hypothesis.

2.3.2 Method

2.3.2.1 Participants

Twenty-nine young adults and twenty-nine older adults were recruited for the study. Data from two young adults and one older adult were discarded due to high levels of performance on the final trial (see below). Data from a further six older adults were discarded due to fatigue leading to incomplete data. The remaining sample contained twenty-seven young adults (age range 18-28, M = 20.63, SD = 3.07) and twenty-two older adults (age range 62-84, M = 72.64, SD = 5.88). All participants in the older adult group attained scores on the MMSE (Folstein et al., 1975) above the cut off of 27 (M = 28.91, SD = 1.11), and were not receiving medication reported to affect cognitive function. The critical issue to control for was the level of French education, therefore participants were asked to report how many years they had spent studying French. No significant differences were found (young adult M = 3.30, SD = 2.23; older adult M = 3.34, SD = 2.33; $t(47) = 0.068$, p = .946), indicating a similar level of expertise between the two age groups. All participants received more than ten years of education, although young adults (M = 15.81, SD = 1.96) reported significantly more years of education than older adults (M = 13.48, SD = 3.23; $t(47) = 2.92$, p = .006). Participants received either course credit or were unpaid volunteers.

2.3.2.2 Materials

A total of 80 French words and their English equivalents were used, all of which were nouns of three to nine letters long. An initial list of 100 medium to high frequency French words was selected from the Brulex database (Content et al., 1990) by one native English speaker and one native French speaker. These words were piloted on eight young adults who had studied French to a maximum of GCSE level (no more than five years) who did not take part in the study. From these 100, I selected 15 words that all pilot participants correctly translated (e.g. maison – house), 15 that some participants were able to translate (e.g. souris – mouse) and the remaining 50 words that no participants were able to translate (e.g. hibou - owl) to ensure sufficient scope for repeated learning trials. In all stages the French word was presented as the cue with the English translation as the required target. A standard recall-judgement-recognition procedure (Hart, 1965) was used to obtain FOK judgements. The recognition task was a four alternative forced choice task, with the same three distracters per target English translation used throughout the procedure (i.e. if the target was ‘mouse’, then the distracters ‘forge’, ‘slate’ and ‘cigar’ were used for each recognition test). The position of targets and distracters was randomised for each recognition test. Distracters were matched to the targets on Kucera-Francis written frequency (targets *M* = 144.02, *SD* = 221.41; distracters *M* = 140.80, *SD* = 224.72) and were not semantically related to each other or the target.

2.3.2.3 Procedure

Participants were tested individually in a single session.

Semantic task. The Semantic FOK task used in the present study was similar to that of Peynircioğlu and Tekcan (2000). The aim here was to assess participants' ability to predict future recognition for non-translated items based purely on conceptual knowledge: participants had not yet been provided with the correct translations for the French words. Prior to any learning, the 80 French words were presented simultaneously, and participants were asked to write down the correct English translation next to the corresponding French word if known. FOKs were only obtained for errors of omission: if participants were unable to give a translation (either correct or incorrect), a dichotomous Yes/No FOK judgement (Souchay et al., 2000; 2007) was made as to whether they would recognise the correct translation for the French word from four options. The four alternative forced choice recognition test was then administered. No time constraint was imposed for recall or recognition.

Episodic feeling of knowing. Immediately after the Semantic trial, the Episodic trials began (Episodic trial 1). All 80 French words and their English translations were presented on a single sheet in three columns for learning: this was the first time during the experimental procedure that participants were shown the correct translations. Words were presented in Calibiri font at font size 14, with French words written in lower case and their English translations written in upper case. The translations were shown for 160 seconds, with only the French word and the English translation provided to encourage episodic word-to-word associations to be made (Kroll & Curley, 1988). Immediately after learning, a cued recall test was given followed by the same four alternative forced choice recognition test, with FOKs obtained for errors of omission only as in the Semantic FOK trial. The learning, recall and recognition were then repeated twice more on all 80 French words (Episodic trials 2 and 3). As in the Semantic trial, no time limit was imposed for recall and recognition. The key difference between the Semantic and Episodic trials is exposure to the correct translations: for the Semantic trial, participants must rely on knowledge held prior to the experimental procedure to inform their FOK judgements, whereas for the Episodic trials information provided during the experimental procedure (i.e. during the learning stage where the correct translations were provided) is used to inform the FOK.

2.3.3 Results

Results are presented in accordance with the two key aims of the study. First, memory and FOK data from the Semantic and first Episodic trial are reported, analysis of which allows examination of the potential age effect on semantic and episodic FOK accuracy. Second, memory and FOK data analysis are presented from the three Episodic trials only, to allow examination of the effect of learning on episodic FOK accuracy in young and older adults. As in Experiment 2.1, a revised Episodic score was calculated for each participant. For this experiment, the strict criterion only was used, as this provided the more conservative delineation between episodic and semantic memory. Any items correctly recalled or recognised by a participant during the semantic stage would not require learning via word-to-word associations, therefore would be less likely to use episodic processes. These items were removed before calculating the participant's Episodic memory and FOK measures as in Experiment 2.1, giving a proportion accuracy for the remaining 'purely episodic' items. This resulted in a mean of 51.00 (*SD* = 9.17) Episodic items for the young adults, and a mean of 56.59 (*SD* = 11.66) Episodic items for the older adults.

2.3.3.1 Semantic vs. Episodic (Trial 1)

Memory

Recall and recognition were each analysed using a 2 (age: young and older adults) x 2 (trial: Semantic and Episodic trial 1) mixed ANOVA. Means and standard deviations are presented in Table 2.7. The effect of age tended towards significance, $F(1,47) = 3.355$, $p = .073$, $\eta^2 = 0.067$, with older adults showing higher recall performance than young adults. Recall was found to be significantly higher in the Semantic than the Episodic trial, $F(1,47) = 18.778$, $p < .001$, $\eta^2 = 0.285$. A significant interaction between age and trial, $F(1,47) = 4.430$, $p = .041$, $\eta^2 = 0.086$, indicated that, for the Semantic trial, older adults showed greater recall performance than young adults, $t(47) = 2.421$, $p = .019$. However, for the Episodic trial, recall performance was comparable across the two age groups, $t(47) = 0.224$, $p = .824$. A main effect of age was also observed in recognition performance, $F(1,47) = 3.805$, $p = .057$, $\eta^2 = 0.075$, with higher recognition accuracy in older adults than young adults. No effect of trial, $F(1,47) = 2.622$, $p = .112$, $\eta^2 = 0.053$, or interaction between age and trial, $F<1$, was present.

Table 2.7: Proportion of items which were correctly recall and recognised across trials for young and older adults

		Young adults (n = 27)	**Older adults (n = 22)**
		M (SD)	***M (SD)***
Proportion of items correctly recalled	Semantic Trial	0.247 (0.165)	0.363 (0.169)
	Episodic Trial 1	0.190 (0.116)	0.198 (0.137)
	Episodic Trial 2	0.450 (0.180)	0.441 (0.238)
	Episodic Trial 3	0.645 (0.211)	0.585 (0.282)
Proportion of items correctly recognised	Semantic Trial	0.638 (0.115)	0.707 (0.146)
	Episodic Trial 1	0.668 (0.166)	0.746 (0.178)
	Episodic Trial 2	0.867 (0.142)	0.875 (0.140)
	Episodic Trial 3	0.940 (0.097)	0.918 (0.124)

In addition to the proportion of items correctly recalled and recognised, the proportion of omission and commission errors at recall were also examined with 2 (age) x 2 (trial) ANOVAs (see Table 2.8). For errors of omission, a main effect of age was observed, $F(1,47) = 10.161$, $p = .003$, $\eta^2 = .178$, with young adults making more omissions than older adults. This leads to a greater number of FOK judgements available to measure metacognitive accuracy for the young adult group. A main effect of task was found, $F(1,47) = 22.489$, $p < .001$, $\eta^2 = .324$, with a larger proportion of omission errors during the Episodic task than the Semantic task. An interaction between age and task was also present, $F(1,47) = 5.339$, $p = .025$, $\eta^2 = .102$. Both age groups gave significantly more omission errors for the Episodic task than the Semantic task (young adults: $t(26) = 2.279$, $p = .031$, older adults: $t(21) = 3.941$, $p = .001$), but while older adults made fewer omissions at the Semantic trial, $t(47) = 3.449$, $p = .001$, a similar proportion of omission were given by both age groups at the Episodic trial, $t(47) = 1.488$, $p = .143$.

Table 2.8: Proportion of omission errors and commission errors given during recall by young and older adults

		Young adults M (SD)	**Older adults M (SD)**
Proportion of omission errors	Semantic Trial	0.716 (0.182)	0.524 (0.208)
	Episodic Trial 1	0.786 (0.116)	0.725 (0.169)
	Episodic Trial 2	0.515 (0.189)	0.481 (0.233)
	Episodic Trial 3	0.325 (0.214)	0.346 (0.268)
Proportion of commission errors	Semantic Trial	0.037 (0.034)	0.113 (0.064)
	Episodic Trial 1	0.024 (0.042)	0.077 (0.095)
	Episodic Trial 2	0.036 (0.061)	0.078 (0.070)
	Episodic Trial 3	0.030 (0.043)	0.069 (0.086)

For commission errors, again a main effects of age was observed, $F(1,47) = 20.103$, $p < .000$, $\eta^2 = .300$. Older adults gave more commission errors, indicating a lower criterion for explicit recall attempts. A main effect of task was also found, $F(1,47) =$

5.723, p = .021, η^2 = .109, with a greater proportion of commission errors during the Semantic trial than the Episodic trial. No interaction was present, $F(1,47)$ = 1.398, p = .243, η^2 = .029.

Metacognition

To examine potential bias in FOK responding, a 2 (age) x 2 (trial) ANOVA was used to compare the proportion of Yes FOK responses given. No effect of age was present, $F<1$, but a main effect of task was found, $F(1,47)$ = 39.105, $p < .001$, η^2 = .454. Participants gave a greater proportion of Yes FOK responses in the Episodic task than the Semantic task. No interaction was present, $F<1$, indicating that young adults and older adults changed responding in a similar way over the trials (Young adults: Semantic M = 0.239, SD = 0.129, Episodic M = 0.460, SD = 0.250; Older adults: Semantic M = 0.262, SD = 0.166, Episodic M = 0.477, SD = 0.314).

A 2 (age) x 2 (trial) x 2 (status: Yes FOK and No FOK) mixed model ANOVA was conducted to examine recognition accuracy for Yes FOKs and No FOKs (see Figure 2.6). Using this method, FOK accuracy would be indicated by higher levels of recognition for Yes FOK responses than for No FOK responses. No main effect of age was present, $F<1$. A main effect of trial was found, $F(1,40)$ = 9.921, p = .003, η^2 = 0.199, with overall recognition higher in the Episodic than the Semantic trial. Critically, a main effect of status was observed, $F(1,40)$ = 125.872, $p < .001$, η^2 = 0.759, whereby Yes FOKs showed greater recognition accuracy than No FOKs, as expected given accurate monitoring of non-recalled information. A significant interaction between trial and status, $F(1,40)$ = 22.317, $p < .001$, η^2 = 0.358, indicated that although recognition accuracy decreased slightly from Semantic to Episodic trials for Yes FOKs, recognition accuracy for No FOKs increased greatly from Semantic to Episodic. The age and status interaction was not significant, $F<1$, suggesting that both groups' subsequent recognition was in keeping with their earlier FOK judgement. The interaction between age and trial type observed, $F(1,40)$ = 4.080, p = .050, η^2 = 0.093, is in keeping with the interaction above, with young adults showing comparable recognition performance over Semantic and

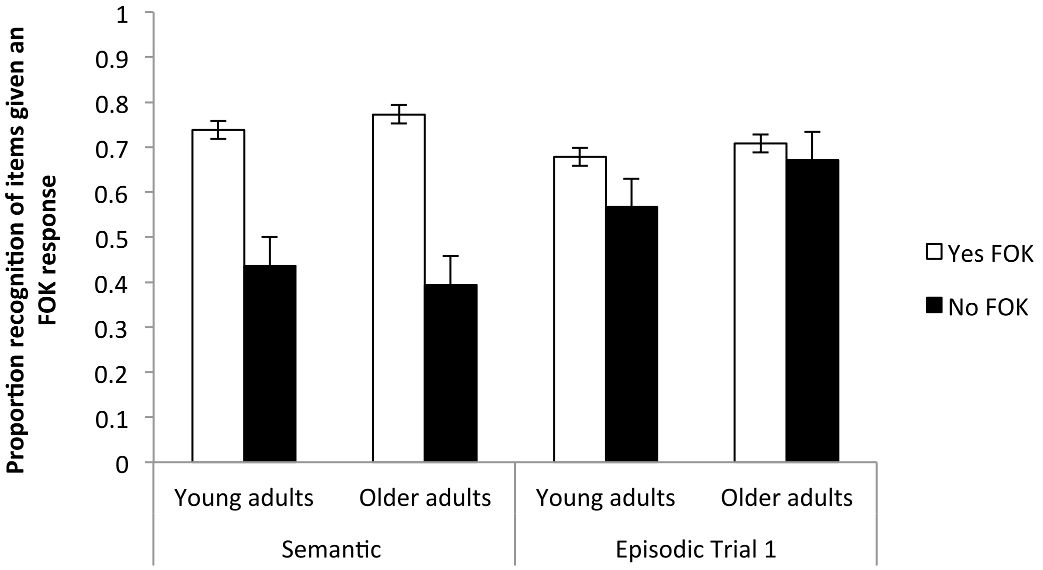


Figure 2.6: Recognition accuracy of items assigned Yes and No FOKs within each age group in the Semantic trial and initial Episodic trial.

Episodic trials and older adults' performance higher in the Episodic condition. No other interactions were detected (trial, age and status: $F<1$).

Using a mixed model 2 (age) x 2 (trial) ANOVA, overall accuracy of FOK predictions showed no main effect of age, $F<1$. A main effect of trial was present, $F(1,47) = 25.505$, $p < .001$, $\eta^2 = 0.352$, with higher gamma scores for the Semantic trial than the Episodic trial (see Table 2.9). No interaction was observed between trial and age, $F(1,47) = 2.334$, $p = .133$, $\eta^2 = 0.047$. To establish if gamma scores were predictive of future recognition, one-sample t-tests were conducted. As can be seen in Table 2.9, for the Semantic trial both young adults', $t(26) = 8.936$, $p < .001$, and older adults', $t(21) = 9.955$, $p < .001$, gamma scores were significantly above zero, indicating FOK judgements were diagnostic of future recognition. For the Episodic trial, although the young adults' gamma scores were again significantly above chance, $t(26) = 3.970$, $p = .001$, the gamma scores for the older adults were not, $t(21) = 0.835$, $p = .413$. For the first Episodic FOK trial, older adults were not able to make accurate predictions of future recognition performance for unrecalled items.

Table 2.9: Gamma correlations for young and older adults for the Semantic trial and Episodic trials 1, 2 and 3

	Young adults (*n = 27*)	**Older adults (*n = 22*)**
	***M* (*SD*)**	***M* (*SD*)**
Semantic Trial gamma score	.512 (.298)**	.591 (.278)**
Episodic Trial 1 gamma score	.243 (.318)**	.089 (.498)
Episodic Trial 2 gamma score	.428 (.458)**	.238 (.573)
Episodic Trial 3 gamma score	.442 (.418)**	.292 (.556)*

*Significant at the 0.05 level

**Significant at the 0.01 level

Pearson's correlations revealed no association between recall accuracy and FOK accuracy for young and older adults within either the Semantic or Episodic tasks (see Table 2.10). This would suggest that differences in recall accuracy are not related to any differences observed in gamma scores.

2.3.3.2 Effect of learning (Episodic trials 1, 2 and 3)

Memory

A 2 (age) x 3 (trial: Episodic trial 1, 2, and 3) mixed model ANOVA was used to examine recall and recognition performance (see Table 2.7). No main effect of age was present, $F<1$. As expected, a main effect of trial on recall performance was observed, $F(2, 94) = 199.957$, $p < .001$, $\eta^2 = 0.810$, with recall increasing over the learning trials. No interaction between age and trial, $F(2,94) = 1.394$, $p = .253$, $\eta^2 = 0.029$, was observed. Regarding recognition performance, again no main effect of age was found, $F<1$, whereas a main effect of trial was observed, $F(2,94) = 130.549$, $p < .001$, $\eta^2 = 0.735$, with recognition accuracy increasing over the trials. However, an interaction between age and trial was observed, $F(2,94) = 6.535$, $p = .002$, $\eta^2 =0.122$, with young adults showing a greater increase in recognition performance than older adults over the three learning trials.

As previously, omission and commission errors were also analysed using 2 (age) x 3 (trial) mixed model ANOVAs (see Table 2.8). Omission errors revealed no main effect of age, *F*<1, but a main effect of trial, $F(1,47) = 222.690$, $p < .001$, $\eta^2 = 0.826$. A significantly lower proportion of omission errors were made with each successive trial. No interaction between age and trial was present, $F(1,47) = 2.164$, $p = .121$, $\eta^2 = .044$. With regards to commission errors, a main effect of age was observed, $F(1,47) = 7.032$, $p = .011$, eta = .130. As previously, older adults were more likely to make commission errors compared to young adults. No effect of trial, *F*<1, or interaction between age and trial, *F*<1, was observed.

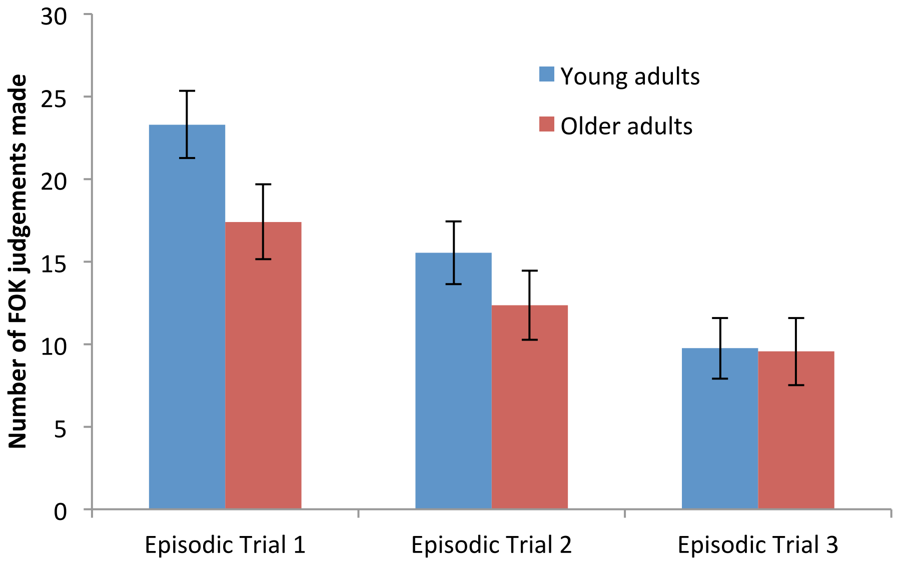


Figure 2.7: The number of items receiving an FOK judgement at each Episodic trial for young and older adults

A further consideration with the design of this task is the number of items on which the accuracy of the FOK is assessed. The greater the number unrecalled items, the more FOK responses can be obtained, giving a more accurate assessment of the ability of the participant to predict future recognition. As learning increases so too does recall, leading to fewer items being eligible for FOK analysis and reducing its stability. The number of items receiving FOK judgements were therefore examined using a 2 (age) x 3 (trial) ANOVA (see Figure 2.7). No main effect of age was present, $F(1,47) = 1.236$, $p = .272$, $\eta^2 = .026$. As expected, a main effect of trial did occur, $F(1,47) = 131.481$, $p < .001$, $\eta^2 = .737$, with increasing learning leading to fewer FOK judgements being made. An interaction between age and trial was also present, $F(1,47) = 9.171$, $p < .001$, $\eta^2 = .163$. While the number of FOK judgements decreases for both age groups over the trials, at Episodic Trial 1 older adults give marginally fewer FOK judgements than young adults, $t(47) = 1.931$, $p = .060$. No age difference is detected at Episodic Trial 2, $t(47) = 1.115$, $p = .270$, or at Episodic Trial 3, $t(47) = 0.071$, $p = .943$.

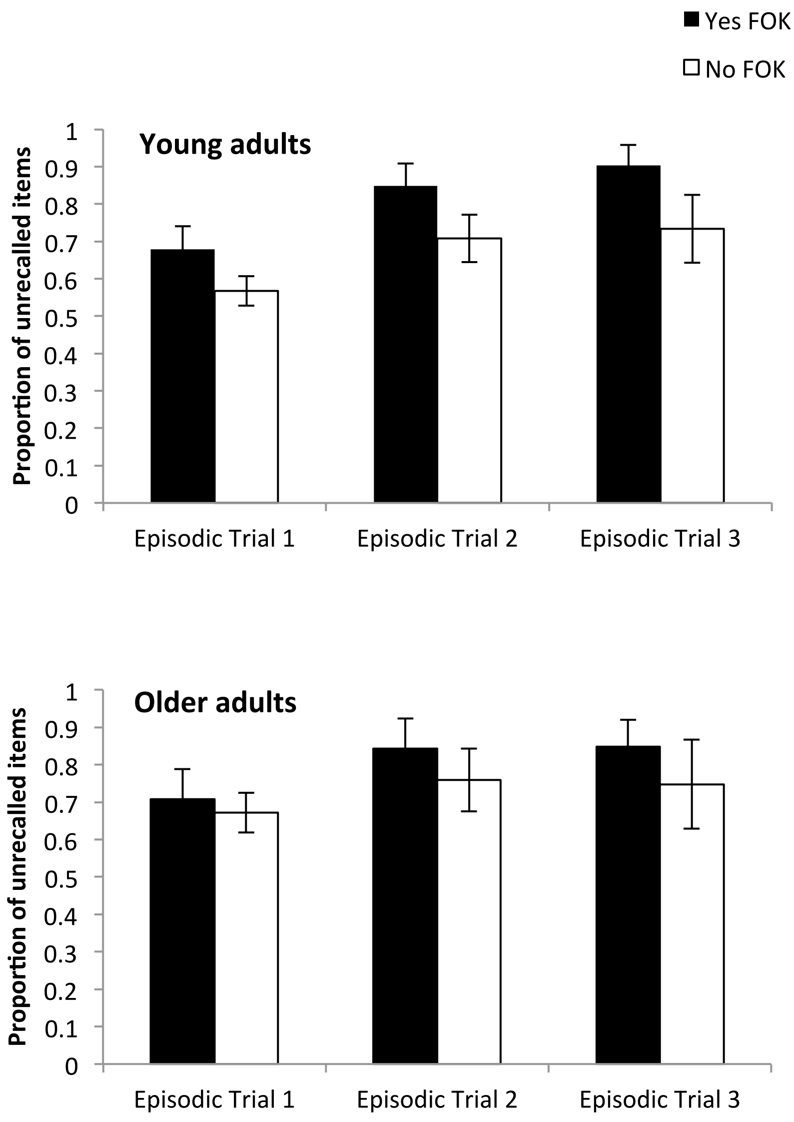


Figure 2.8: Proportion of items subsequently recognised for Yes FOKs and No FOKs in young and older adults

Metacognition

Bias towards Yes FOK responding was examined with a 2 (age) x 3 (trial) ANOVA. No effect of age was present, $F<1$, as previously. A main effect of trial was observed, $F(2,94) = 28.635$, $p < .001$, $\eta^2 = .379$, with the proportion of Yes FOKs increasing over the learning trials. No interaction was present, $F(2,94) = 1.944$, $p = .149$, $\eta^2 = .040$, again suggesting that young and older adults were similarly affected by the repeated learning trials (Young adults: Trial 1 $M = 0.460$, $SD = 0.250$, Trial 2 $M = 0.701$, $SD = 0.290$, Trial 3 $M = 0.810$, $SD = 0.244$; Older adults: Trial 1 $M = 0.477$, $SD = 0.314$, Trial 2 $M = 0.594$, $SD = 0.363$, Trial 3 $M = 0.694$, $SD = 0.403$).

A 2 (age) x 3 (trial) x 2 (status) mixed model ANOVA was conducted to examine recognition accuracy for Yes and No FOKs (Figure 2.8). No age effect was present, $F(1,17) = 2.622$, $p = .124$, $\eta^2 = 0.124$. A main effect of trial was observed, $F(2,34) = 19.332$, $p < .001$, $\eta^2 = 0.532$, with recognition increasing over trials. However, no main effect of status, $F(1,17) = 2.524$, $p = .131$, $\eta^2 = 0.129$, was found. All interactions failed to reach significance (trial and age: $F<1$; status and age: $F(1,17) = 2.049$, $p = .170$, $\eta^2 = 0.170$; trial and status: $F<1$; trial, status and age: $F<1$).

A 2 (age) x 3 (trial) mixed model ANOVA on adjusted gamma scores showed a tendency towards significance for age, $F(1,47) = 2.908$, $p = .095$, $\eta^2 = 0.058$, with young adults gamma scores greater than those of the older adults. A main effect of trial was also observed, $F(2,94) = 3.457$, $p = .036$, $\eta^2 = 0.069$, with increasing accuracy over the learning trials (see Table 2.9), indicating that participants became more accurate in their judgements with increased task experience. No interaction between trial and age was observed, $F<1$. The predictive accuracy of gamma scores for each age group at each learning trial was examined. For the young adults, gamma scores at Episodic trial 1, $t(26) = 3.970$, $p = .001$, Episodic trial 2, $t(26) = 4.854$, $p < .001$, and Episodic trial 3, $t(26) = 5.502$, $p < .001$, were all significantly above chance, indicating that young adults were accurate in their predictions for future recognition. As described above, the older adult group were not above chance at the Episodic trial 1 FOK judgement, $t(21) = 0.835$, $p = .413$. By Episodic trial 2 there is a trend towards significance, $t(21) = 1.951$, $p = .065$; however, it is not until Episodic trial 3 that older adults showed predictive accuracy for recognition performance, $t(21) = 2.463$, $p = .023$.

As with the Semantic-Episodic analysis, Pearson's correlations were examined to establish whether recall accuracy was related to FOK accuracy. As can be seen in Table 2.10, no relationship was observed for young or older adults within any of the Episodic trials.

Table 2.10: Pearson's correlations between recall accuracy and gamma correlations for young and older adults at each trial

	Young adults (N = 27)		Older adults (N = 22)	
	r	***p***	***r***	***p***
Semantic	-.132	.511	.034	.880
Episodic Trial 1	.276	.163	.296	.181
Episodic Trial 2	.221	.267	-.014	.949
Episodic Trial 3	.270	.172	-.007	.975

2.3.4 Discussion

The aim of Experiment 2.2 was to further test the presence of an age-related deficit in episodic FOK accuracy. A similar paradigm was used to Experiment 2.1, which allowed assessment of both semantic and episodic memory within the same participants while controlling for the effects of task characteristics. In addition to replicating the basic paradigm with a novel language learning task, the present study also allowed assessment of the effect of repeated learning trials on episodic FOK accuracy. With regards to the semantic episodic distinction, results further support previous findings that older adults are capable of making accurate Semantic FOK judgements (e.g. Butterfield et al., 1988), with gamma accuracy at a similar level to that of young adults. However, the findings for Episodic FOK are less straightforward. No influence of age can be detected when comparing Gamma correlations between young and older adults. Yet when examining Gamma accuracy within each age group, older adults are at chance levels of performance within the Episodic trial, whereas young adults are reliably above chance for Episodic FOK accuracy. Despite no age effect or interaction between age and task, older adults do in fact show a deficit in being able to predict future recognition of unrecalled items. Likewise, when considering the learning aspect of the data, it would appear that both young and older adults improve their FOK accuracy as the number of learning trials increase, with no age effects being present. Yet when looking at individual Gamma correlations for each group, it is not until Episodic trial 3 that the older adults are able to predict future performance with above chance accuracy, whereas young adults are able to do so in all trials. Thus, despite the main analysis indicating no age deficit in relation to Episodic FOK accuracy, there are some indications in the data that this result should be interpreted with caution.

Likewise, the explanatory value of the Memory Constraint Hypothesis (Hertzog et al., 2010) is somewhat limited. No effects of age are observed on the Episodic memory measures, and both age groups show similar memory improvements due to the repeated learning. In addition, no correlations were observed between recall and FOK accuracy. The MCH would therefore predict that the information on which young and older adults base their FOK judgements will be of a similar quality, thus a similar level of FOK accuracy should be observed. As described above, this is indeed the case, with no age effects or interactions being shown. However, the chance performance of older adults for the first two Episodic trials cannot be ignored, suggesting that memory performance may not be the sole explanatory factor of metacognitive accuracy in this experiment.

As in Experiment 2.1, no age effect was shown on Episodic memory performance. This is somewhat surprising given the overwhelming evidence of episodic memory deficits in older adults typically reported in the literature. This pattern of results in the current study could potentially be explained by the use of the language learning paradigm. As previously mentioned, while standard word-pair learning can only be achieved via episodic associative memory processes, language learning will involve the formation of links with semantic memory as well as episodic memory processes. The Revised Hierarchical Model (RHM) of bilingual memory (Kroll & Stewart, 1994) proposes that novice bilinguals initially rely on the episodic word-to-word associations and, as proficiency increases, shift to semantic-based word-to-concept associations. This shift from episodic to semantic based processing has been observed in young and older adults (Service & Craik, 1993), and although the timing of the shift was not specified in the original model, early data indicated that episodic learning still occurred up to two years after commencement of learning (Kroll & Curley, 1988). However, further studies have shown semantic-based processes being recruited in word learning even within a single study session (Altarriba & Mathis, 1997; Papagnio et al., 1991).

The present study was designed to encourage the use of episodic based learning processes by presenting the word forms at learning without any contextualising information, and by requiring participants to translate from the second language to the first language which, according to RHM utilises episodic, lexical level links (Sholl et al., 1995). Despite this it appears that semantic processes were still recruited during learning (consistent with the findings of Altarriba & Mathis, 1997). Combined with the observation of superior semantic memory performance in the older adults, it is reasonable to suggest that the lack of episodic memory deficit within the current study is due to involvement of semantic processes during learning. Although speculative, the possible contribution of semantic memory processes to episodic tasks of similar design to the current study warrants further investigation.

The greater level of semantic memory in the older adult group raises the issue of proficiency as a possible confound. To date, only three experiments have examined the role of expertise in FOK accuracy. Roberts and Rhodes (1989) showed no effect of self-rated expertise in the accuracy of FOKs for general knowledge information, a finding repeated by Marquié and Huet (2000) looking at self-efficacy beliefs for computer knowledge in young and older adults. Of direct interest to the current study, Peynircioğlu and Tekcan (2000) objectively measured language expertise and its effect on FOK ratings. They observed an increase in the magnitude of FOK ratings with increased expertise. However, the accuracy of FOK ratings showed no effect of expertise. Although participants with greater proficiency showed higher memory performance and were more confident in their ratings of future recognition, this was not predictive of their actual recognition performance. Taken together, these studies suggest that language expertise did not influence the accuracy of the FOK judgements within the present study.

2.4 General discussion

The aim of the present chapter was to explore the effect of age on semantic and episodic FOK accuracy. Previous research has been somewhat unequivocal regarding whether or not older adults show a selective deficit in being able to predict future recognition of unrecalled episodic items compared to semantic items. As shown by studies into the effects of learning on episodic FOK accuracy (Carroll & Simington, 1986; Carroll & Nelson, 1993; Nelson et al., 1982) this could be due to the use of between subjects designs being prone to the influence of individual differences, particularly where subjective decisions are concerned. Two experiments were therefore conducted to measure semantic and episodic accuracy within the same participants, and with the same task characteristics, in order to limit as many extraneous sources of bias as possible. Results from Experiment 2.1 show a clear age effect on episodic FOK accuracy only. While both young and older adults are able to accurately predict recognition for semantic items, older adults are significantly below the levels of performance shown by young adults for episodic items. Experiment 2.2 is somewhat less definite, with the suggestion of an age effect as evidenced by older adults' Gamma scores being at chance level, and yet no main effects of age or interactions with age being detectable within the primary analysis. Taken together, these experiments do support previous findings of age-related impairments in episodic FOK accuracy (Perrotin et al., 2006; Souchay et al., 2000, 2007; Thomas et al., 2011), however the ability to detect these impairments may be heavily influenced by the task used.

The pattern of results observed in the current chapter support the inferential hypothesis of selective impairments in FOK accuracy in ageing. Taken together, data from neuroimaging (Kikyo & Miyashita, 2004; Kikyo et al., 2002; Maril et al., 2003; Reggev et al., 2011), studies in ageing, executive functioning and FOK (Perrotin et al., 2006; Perrotin, Tournelle, & Isingrini, 2008; Souchay et al., 2000; Souchay & Isingrini, 2004), and ageing and recollection (Bunce, 2003; Perfect & Dasgupta, 1997; Souchay et al., 2007) all suggest that older adults would show preserved semantic FOK accuracy, while episodic FOK accuracy would be impaired. Experiments 2.1 and 2.2 both exhibited this pattern of results to some extent, either with the presence of a clear age effect or with older adults being at chance in their predictive accuracy for episodic times only. The MCH on the other hand is unable to account for the findings of the experiments presented within this chapter. In both instances older adults' episodic memory accuracy is equivalent if not superior to that of young adults, and yet FOK accuracy as measured by Gamma correlations is significantly lower for older adults than young adults, or at chance levels. The MCH would predict that where episodic memory accuracy is at similar levels, episodic FOK accuracy would also be at similar levels. Similarly, the MCH would assume a relationship would exist between recall accuracy and FOK accuracy, at least within the older adult group if encoding drives previously observed FOK deficits. The data throughout this Chapter fail to find any such association. The MCH does, however, provide a good explanation for the effect of learning on FOK accuracy. As repeated exposures to the cue-target pairing increase, so too does the quality of the information on which the FOK is based, leading to increased accuracy.

In the language learning paradigm, the Semantic FOK task can utilise both cue familiarity and target accessibility to inform FOK judgements. Only those items for which the translation has been learned at some point in the past will look familiar. However, for the initial Episodic trial, all items which are correctly recognised at the Semantic trial have been removed from the analysis. This leaves only items which participants do not know the meaning of, which could essentially be viewed as nonwords. Prior to learning, it is therefore reasonable to assume that these items will all have a similar but low level of familiarity. During learning however, the familiarity of each of the cue words will change (Reder & Ritter, 1992). As noted previously, young and older adults have been observed to have similar selection strategies when learning an unfamiliar language (Price et al., 2010), therefore changes in cue familiarity should be equivalent between the two age groups. Item differences in cue familiarity can be used by young and older adults to inform their FOK judgements, thereby reducing the effect of impaired target accessibility in the older adult group. The increased diagnostic value of cue familiarity could consequently lead to a reduced deficit in episodic FOK accuracy in the older adult group, whereby performance is not significantly above chance but is not significantly below that of the young adults, as observed in Experiment 2.2.

One consideration when interpreting the FOK accuracy results for the current chapter is the presence of age differences in the proportion of commission errors observed in both studies. If we adopt a memory strength view of recall, recognition and FOK, the presence of higher commission errors for older adults implies a liberal criterion for reporting free recall responses. This will subsequently reduce the range of memory strength of the remaining items on which the FOK is based: any high memory strength items will be reported, leaving only low memory strength items for FOK inspection. Young adults meanwhile, by having a stricter recall criterion, will have some high memory strength items available for FOK. This greater difference in memory strength will consequently make it easier to discriminate items which will subsequently be recognised from those items which will not. Older adults, with a smaller range of memory strength items, will have a more difficult task to discriminate recognisable items from nonrecognisable items, thus reducing their FOK accuracy. This argument would therefore suggest that age differences in FOK accuracy are due to differences in task difficulty rather than an age effect. However, this proposal is somewhat called into question by the increased FOK accuracy observed in Experiment 2.2 with repeated learning trials. Recall increases over the repeated learning trials, thus reducing the number of items on which the FOK can be based and, presumably, the range of memory strength of those remaining items. And yet, FOK accuracy increase over the trials rather than decreases. Thus, while differences in task difficulty may have contributed to the observed age differences in FOK accuracy, further examination of the role of reduced ranges of memory strength is needed to clarify its contribution.

It is worth noting that FOK accuracy tends to (but not always) be higher in semantic tasks than episodic tasks independent of the effects of ageing. For example, Perfect & Hollins (1996) observed average gamma correlations of 0.37 for a general knowledge task while for an episodic task correlations were at 0.10. While neuroimaging has observed that semantic and episodic FOK do involve unique as well as shared regions of activation (Reggev et al., 2011), consideration of what FOKs are based on also implies the two memory tasks may lead to different levels of FOK accuracy, as suggested by Perfect & Hollins (1999). The FOK is based on a combination of partial information, which includes the familiarity of the cue and details partially recalled about the target. For a semantic task, the associative link between the cue and the target works in both directions; there are typically few unique answers to a question and few unique questions that could prompt a specific target. In episodic tasks, however, the link is only diagnostic in a single direction, from the cue to the target. This therefore reduces the informativeness of the familiarity of the cue. With regards to partial information retrieved, the semantic cue will activate related information which may also aid access to the correct target later, whereas for an episodic cue this related information is unlikely to be useful for recognising the target. Thus both cue familiarity and partial information are more diagnostic in semantic tasks than episodic tasks, leading to greater predictive accuracy (Perfect & Hollins, 1999).

The lack of an age-associated episodic memory deficit could be due to the amount of environmental support available. The environmental support hypothesis (Craik, 1983, 1986) proposes that older adults are unable to self-initiate processes which would aid encoding and retrieval. By providing external cues and thus reducing reliance on internal processes, the age-related deficit in memory accuracy can be reduced or even eliminated (e.g. Naveh-Benjamin, Craik, & Ben-Shaul, 2002). The general knowledge task used in Experiment 2.1 provides an elaborate cue for the target; the framing of the question and the wording used provide more support than using a single word as a prompt as is typically used in episodic memory tasks. Likewise, the French learning paradigm used in Experiment 2.2 could be recruiting semantic memory processes (Altarriba & Mathis, 1997; Papagno, Valentine, & Baddeley, 1991) which again would provide more support than a typical episodic FOK task. This increased support inherent within the paradigms could subsequently lead to higher episodic memory accuracy than is typically observed. The role of environmental support at retrieval will be considered in greater detail in Experiments 3.3 and 3.4.

One potential drawback of the paradigms used within the current chapter is the structure of the learning phase for each experiment. Participants were provided with all items during learning, items that they had successfully recalled and items which required encoding. This requires participants to then select which of the items presented they need to allocate resources to for successful retrieval at test. Numerous studies have observed an age-related deficit in self-initiated strategy use during encoding (e.g. Dunlosky & Connor, 1997; Murphy, Sanders, Gabriesheski, & Schmitt, 1981; Murphy, Schmitt, Caruso, & Sanders, 1987). By providing all items at learning, this could be a disadvantage for the older adult group, who may not initiate or employ an appropriate selection strategy to enable maximal learning. Based on the results obtained by the current studies, however, it would not appear that older adults were in any way adversely affected by the structure of the learning phase. Episodic memory performance is at a similar level to that of young adults, suggesting suitable strategy selection and implementation to allow successful retrieval. Furthermore, Price, Hertzog, & Dunlosky (2010) have shown that during self-regulated language learning young and older adults select translations to study in a similar fashion, suggesting that during Experiment 2.2 in particular, older adults were as able as young adults to select the appropriate items for study. It may also be argued that the learning task is at a somewhat lower level of difficulty for the older adults due to their superior semantic memory in both tasks. As more items are correctly retrieved prior to learning, the number that need to be encoded is subsequently reduced. This possible discrepancy in the difficulty of the learning task does not however impact on the conclusions that can be drawn from the present studies. If indeed difficulty is inadvertently reduced for the older adult group, this has ultimately matched our participant groups on episodic memory ability, thus removing differences in memory performance as a possible influence on FOK accuracy.

The present paradigm also raises the issue of which items to class as episodic, as all items are presented at all stages. In Experiment 2.1 a liberal criterion was used initially, with episodic items being identified as those which were not correctly recalled during the semantic task. However this allows items which may be semantic in nature but of insufficient strength to be retrieved for recall to also be included in the measurement of episodic memory. The strict criterion gives the most 'pure' estimate of episodic memory available within the current paradigm by also removing items which are successfully recognised during the semantic task. This criterion only was therefore used in Experiment 2.2. As already discussed, one potential limitation of Experiment 2.2 is the possibility that semantic memory processes are recruited during the episodic trials. Although the task was designed to emphasise episodic processes as much as possible by encouraging learning by word-to-word associations (Sholl, Sankaranarayanan, & Kroll, 1995), it is possible for semantic processes to be recruited to aid episodic learning even within a single learning trial (Altarriba & Mathis, 1997; Papagno et al., 1991).

The above findings relate to the language learning literature and are of direct concern to Experiment 2.2, however the potential influence of semantic processes in Experiment 2.1 cannot be ignored. It may be that the paradigm adopted here encourages participants to use their semantic knowledge to aid episodic learning, due to the identical nature of the semantic and episodic task. The interrelated nature of these two types of memory makes it more difficult to disentangle the effects of ageing on each process individually, an issue which is largely ignored when considering each memory in a separate task. Nonetheless, the findings of the recollection analysis in Experiment 2.1 support the assumption that we have separated semantic and episodic memory processes to a sufficient extent to allow us to establish any differences in metacognitive processes associated with each type of memory. We can therefore be confident in the findings obtained by the paradigm, and the conclusions drawn from them.

A final consideration of the MCH is warranted due to the results of the current chapter being in contrast with those predicted by this hypothesis. Although the MCH does not propose a direct link between memory and FOK, it does propose that higher memory performance leads to better quality information being available for inferential processes such as the FOK to work with. This therefore leads to more accurate inferential processing, leading to high memory performance being associated with higher FOK accuracy. When examining the previous literature supporting an age effect on episodic FOK accuracy, the results of Souchay et al. (2000; 2007) and Perrotin et al. (2006) indicate that, in addition to FOK deficits, older adults exhibited significantly lower recall and recognition performance than young adults. Participants in the Thomas et al. (2011) study, where an age effect was also found on the episodic FOK accuracy, showed an age-related reduction in recall performance only; recognition accuracy was not impaired. Conversely, the results of MacLaverty and Hertzog (2009) showed no effect of age on episodic FOK accuracy. However, an age-related decline in recall and recognition performance was observed. These studies represent a mixed picture of the association between measures of memory performance and episodic FOK accuracy, with no clear pattern of whether impairment in one necessarily results in impairment of the other. Therefore although the MCH can to some extent explain why some studies report an age effect on episodic FOK accuracy, other cognitive or memory factors might interfere or contribute to this relationship, thus indicating that alternative explanations also need to be considered.

In sum, the experiments presented in the current chapter support the hypothesis that older adults show a selective deficit in episodic FOK accuracy, while semantic FOK accuracy is preserved.

Chapter 3. Feeling of Knowing Accuracy and Manipulations of Retention and Retrieval

3.1 General introduction

The purpose of the current chapter is to further explore the potential link between the memory and its influence on FOK magnitude and resolution. As outlined in Chapter 1, the precise relationship between memory and FOK is not known, and manipulations of memory performance do not always result in similar effects on FOK accuracy. Thus far, examinations of this relationship have focused on manipulations at encoding. The present chapter aims to examine the effect of manipulations at retention and retrieval. In addition, a more sensitive FOK judgement will be obtained.

The measurement of the FOK used in Experiments 2.1 and 2.2 was a simple Yes/No judgement as in the original Hart (1965) study; the participant reports that they believe they will correctly recognise the answer or that they will not recognise it. This could be considered as restricting the conception and interpretation of the FOK to a state that is either experienced or not, whereas it may be better reflected as being on a continuum from Yes to No. Empirically, by constraining participants to assign judgements to one of two categories can result in a loss of information (Masson & Rotello, 2009). One example of this comes from Koriat (1975), who observed that when people guessed responses and were correct, FOK ratings were higher than when they guessed responses and were incorrect. Despite participants believing they were guessing the answers, and would therefore have given a No FOK response to both in the binary version, the FOK rating was able to distinguish which was more likely to be correct. Thus by allowing participants to make finer grained judgements it will presumably provide more information about how well they are able to discriminate between items (see Goldsmith, Koriat, & Weinberg-Eliezer, 2002, for a similar line of argument with regards to memory accuracy).

By using a rating scale for FOK responding, it is possible to detect changes in the magnitude of the FOK as well as the accuracy of FOK, which may provide more information as to what effect the memory manipulations are having on FOK. Two classic manipulations will be explored. First, the effect of a delay placed at various points within the standard FOK paradigm (Experiments 3.1 and 3.2). Second, the amount of environmental support provided at retrieval (Experiments 3.3 and 3.4). The potential effect of ageing on FOK performance will continue to be examined within each of these manipulations.

3.2 Experiment 3.1: Delay placement and FOK – between subjects

3.2.1 Introduction

As outlined in Chapter 1, metacognitive judgements are thought to depend on one of two major sources of information: heuristics and deliberate inference (Kelley & Jacoby, 1996; Koriat, 1997). Judgements based primarily on heuristics, known as experience based judgements, are a result of the online processing of information, the mnemonic cues generated during encoding or retrieval of items that give rise to the subjective experiences associated with memory. In contrast, judgements based on deliberate inference (theory based judgements) rely on the application of beliefs and theories about how one's memory works and the impact of task characteristics on memory performance.

One of the earliest emerging beliefs about memory function is the effect of a delay. Children as young as four years old are aware that as the time between encoding and retrieval increases, memory performance declines (Lyon & Flavell, 1993). Koriat, Bjork, Sheffer, & Bar (2004) demonstrated the impact that reliance on theory versus experience based processing can have on the magnitude of metacognitive judgements by examining global Judgements of Learning (JOLs) for variable delays. Global JOLs are predictions of what percentage of items the participant thinks they will be able to correctly recall at test (Nelson & Narens, 1990). As the JOL occurs at the encoding stage, it is possible to isolate the contributions of experience and theory based information when predicting the effects of a delay by eliciting the JOL either prior to or after encoding. By asking participants to make the JOL prior to encoding, only theory based judgements can be made, as the participant has no knowledge of the items to be learned. Consequently, predictions of future recall should be sensitive to the anticipated delay between encoding and recall as individuals will apply what they know about forgetting and memory to their judgements. Once encoding has occurred, the individual can use an online assessment of cues such as perceptual and retrieval fluency in order to predict future performance, giving experience based JOLs.

In this instance, the delay would have no impact on the JOLs, as the JOL would primarily be influenced by the mnemonic cues available at encoding which remain the same regardless of the anticipated delay. Contrary to these predictions, Koriat observed that the effect of delay on theory based judgements was only detectable when comparative judgements were made. With a between subjects design, JOLs were insensitive to delays of one day, one week or even one year: predictions of future recall were the same irrespective of the upcoming delay. It was only by asking participants to make multiple judgements that the expected effects of delay were observed, with JOLs reducing as delay increased. Although participants were aware of the effects of forgetting, they did not spontaneously apply this information unless the task implicitly activated the idea.

The FOK paradigm provides another avenue for investigating the effects of delay and the successful recruitment of experience and theory based judgements. Somewhat surprisingly, the effect of delay on FOK has received little attention. Schacter (1983, Experiment 1) examined the effect of a one week delay on FOK responses. Following encoding of cue-target word pairs, participants were given a recall test for half of the materials immediately after learning and were asked to give Yes-No FOK judgements. One week later, participants were tested on the remainder of the encoded materials, and again asked to make Yes-No FOK judgements. As expected, recall and recognition performance declined over the week long delay. In terms of FOK, participants were less likely to give Yes FOK predictions following the delay than when asked immediately, suggesting a reduction in the strength of the FOK experience. However, when examining the proportions of Yes and No FOKs correctly recognised, no interaction with delay was found: the likelihood of correctly recognising items given Yes or No FOK's was the same whether participants were tested immediately after learning or one week later. Janowsky, Shimamura, & Squire (1989) also examined the effect of delay on FOK judgements in patients with frontal lobe lesions and healthy controls. In this experiment, participants were tested on all items both immediately after learning and after a one to three day delay. Magnitude of FOK responding is not, unfortunately, reported, however a significant effect of delay on FOK resolution was observed. Both control subjects and patients showed a decrease in FOK accuracy when tested at 1-3 days following encoding as opposed to immediately.

Due to the nature of the standard FOK paradigm (Hart, 1965), the FOK judgement is made following a failed recall attempt for an *immediate* recognition test, thus is only based on subjective experience driven by the partial information and mnemonic cues available at the time of making the prediction of future memory performance (Koriat, 1993). By placing the delay between encoding and recall, as in the above two studies, the FOK judgement is made immediately prior to the recognition test. Irrespective of the length of delay, online assessment of the partial information available is all that is necessary to predict recognition for a test which will immediately follow. To mirror the JOL effect of varying reliance on experience and theory based information within the FOK paradigm, the placement of the delay would need to be between the FOK judgement and the recognition test. In this instance, the online assessment of partial information would be insufficient to predict future recognition, as the delay will lead to forgetting and therefore affect retrieval of items at the recognition test. The individual would need to use their knowledge of the effects of forgetting over time to account for the anticipated delay and the subsequent reduction in retrievability of items at test. It would then be expected that the magnitude of the FOK response made at a certain time interval prior to the recognition test would be lower than an FOK prediction made immediately before the recognition test.

This pattern of judgement-delay-recognition has been utilised previously by MacLaverty & Hertzog (2009), who were trying to establish if the delayed JOL effect was also present for FOKs. With JOLs, a short delay between encoding and eliciting the JOL significantly improves the accuracy of the JOL (Nelson & Dunlosky, 1991). MacLaverty and Hertzog (2009) attempted to establish if a short delay at various stages of the FOK paradigm could improve the accuracy of the FOK. To mirror the delayed JOL effect, only short delays of three minutes were used, consequently minimal forgetting will have occurred. No difference in accuracy was observed between immediate and delayed FOK judgements, suggesting that no delayed FOK effect occurs. Thus although the placement of delay in the MacLaverty and Hertzog study is ideal to elicit theory based FOK judgements, the results of the study are not able to establish whether the belief about the effect of delay has been used to inform the FOK judgement. The present study is more aptly described as an investigation of the effects of delay on FOK, rather than a delayed FOK study.

It is important to note that Koriat et al.'s, (2004) assessment of the contributions of theory and experience based judgements to JOLs focuses on the effect of these heuristics on the magnitude of JOL predictions. Consequently, the results of this study can only be directly applied to prediction of the effects of delay on the magnitude on FOK judgements. Changes in the accuracy of the FOK are also of interest. The accessibility account (Koriat, 1993) and the memory constraint hypothesis (MCH: Hertzog, Dunlosky, & Sinclair, 2010) are two main theories as to why FOKs are accurate at predicting future recognition, and each makes slightly different predictions about the effect of a delay on the FOK. As previously explained (Chapter 1), the MCH proposes that FOK resolution is dependent on the quality of the original encoding. Placement of a delay between the FOK and the recognition task should subsequently have no effect on FOK accuracy, as no manipulation of encoding quality has occurred. The accessibility account emphasises the role of retrieval processes, with FOK being dependent on the quality and quantity of the partial information retrieved during the failed recall attempt. When a delay is anticipated, the information accessed during failed retrieval is not sufficient to predict future recognition, thus the application of theory based information is necessary. However, the quality of this theory based information is unknown. If the information is accurate, no effect of delay on FOK accuracy will be found. If the information is not accurate, and the individual under or overestimates the impact of the delay, then the accuracy of the FOK will reduce.

To help assess whether individuals are accurate in their application of theory based judgements, and thus test the predictions of the MCH and accessibility accounts, the current study will examine two placements of a two hour delay in addition to a control condition where no delay is used. As previously described, to encourage reliance on theory based judgements, a delay will be placed after the FOK judgement and before the recognition test. Participants were informed of the placement of the delay prior to testing. When compared to the control condition, it would be expected that recall accuracy would be at similar levels and FOK magnitude would be reduced for participants anticipating the delay. The second delay group will experience the same delay placed between encoding and recall. This group are then able to make experience based FOK judgements only; the recognition test will follow immediately after the FOK and so online assessment of the partial information retrievable after the delay is sufficient. Comparison of the two delay groups should reveal whether the application of theory based judgements was successful or not and, if not, allow us to establish whether the accessibility account or MCH is more capable of explaining the results obtained. If the theory based judgements are not accurate, the accessibility account would predict that the FOK accuracy will be reduced when FOK judgements are made prior to the delay, whereas the MCH would predict the same level of FOK accuracy irrespective of delay placement.

The current study also allows further assessment of the link between recollection and FOK. As described in Chapter 2, associations between recollection and FOK have repeatedly been demonstrated (e.g. Hicks & Marsh, 2002; Thomas, Bulevich, & Dubois, 2011), leading to the suggestion that recollection could contribute to the mnemonic cues that form the basis of the FOK (Souchay et al., 2007). The delay paradigm provides an additional opportunity to clarify the circumstances in which recollection and FOK are related, and also the effects of delay on recollection itself. Previous studies into forgetting and recollection have typically examined either short delays measured in terms of intervening items (e.g. Yonelinas & Levy, 2002) or longer delays of at least one day (e.g. Knowlton & Squire, 1995), with findings showing that at short delays familiarity declines while recollection is preserved, whereas longer delays lead to decreases in both recollection and familiarity. Although some previous research has examined recollection at an intermediate delay (e.g. Hockley & Consoli, 1999), no immediate measure of recollection was obtained, thus it is not possible to say conclusively whether or not the levels of recollection reported are the same as when tested immediately or whether slight deterioration has occurred.

Irrespective of whether the two hour delay results in a reduction in recollection or not compared to the control condition, for both placements of the delay the effect on recollection will be the same. Whether the delay occurs before or after the FOK prediction, the Remember/Know judgements will be obtained following the same time interval, thus in both experimental conditions we would expect the levels of recollection and familiarity to show similar effects. As already described, the same cannot be said for the FOK judgement. If the FOK is primarily experience based, then it would be expected that FOKs made prior to the delay will be less accurate than those made following the delay. However, if theory based judgements are also employed to make the FOK, this potential deficit could be removed. Subsequently, it is unclear whether the relationship between recollection and FOK will be the same for the two experimental conditions or whether slight differences will be detected.

3.2.2 Method

3.2.2.1 Participants

Sixty eight undergraduate students (age range 18 to 30, *M* = 21.39, *SD* = 3.40, 16 male) from the University of Leeds participated in the study in return for course credit. Participants were tested in groups of between one and eight.

3.2.2.2 Materials

A list of 100 concrete nouns was selected from the MRC Psycholinguistics database (Coltheart, 1981). Words were between four and eight letters long, and had an average concreteness rating of 595.91 (*SD* = 24.49, taken from a corpus ranging from 100 to 700). To create novel associations for learning, the nouns were randomly combined to create 50 cue-target word pairs that were not semantically related (e.g. sail – alley). Distracters for the recognition task were other possible targets, i.e. distracters were items which had been paired with a different cue word during learning (Patterson & Hertzog, 2010). Each target word was used as a distracter an equal number of times.

3.2.2.3 General Procedure

Participants were assigned to one of three conditions: two experimental conditions, FOK Before and FOK After, and a Control condition (see Figure 3.1). A basic three stage procedure was followed by all participants, beginning with a learning stage, followed by a cued recall and FOK, then finally a recognition and recollection stage. The two experimental conditions also incorporated a two hour delay, either between cued recall & FOK and recognition & recollection (FOK Before the two hour delay), or between learning and cued recall & FOK (FOK After the two hour delay). Participants were informed prior to testing where the delay in the task would occur, thus encouraging them to use theory based information.

Learning. The 50 cue-target pairs were presented one at a time via Powerpoint, with the cue word presented in lower case above the target word which was presented in upper case. Word pairs were presented in Calibri font size 60, for five seconds each with a one second fixation between each pair. Participants were asked to learn the word pairs for a future memory test.

Cued Recall & FOK. An answer booklet was used to record recall and FOK responses. The task was presented in three columns, with the first column containing the cue words. Presentation order of cues differed from learning. The second column was left blank for recall responses. Participants were asked not to guess at this stage. The final column was for recording FOK responses. All items, whether recall was attempted or not, were given an FOK rating. The FOK in the current study was described as a 'Likelihood Rating', i.e. the likelihood of correctly recognising the required target from four options. Participants were informed that the likelihood ratings could range from 25%, a chance or guess response, to 100%, complete certainty that they would correctly identify the target from four options (MacLaverty & Hertzog, 2009). Full use of the scale was encouraged, and participants were advised to give any number between 25 and 100 as their Likelihood Rating[2]. The recall and FOK stage was not timed, and the next stage of the experiment began when the last participant in the group had completed all recall and FOK responses.

Recognition & Recollection. The recognition task was presented via Powerpoint, with participants recording recognition and recollection responses in the answer booklet. Cue words were displayed at the top of the screen in lower case, with the four possible targets presented in upper case, using Calibri front size 48. Targets were presented in a 2 x 2 grid formation, with each location assigned as A (top left), B (top right), C (bottom left) and D (bottom right). The location of the correct target was randomly assigned across trials. Participants were asked to circle the corresponding letter in their booklets, and were informed that a response was required for every item. Immediately after selection, participants were asked to categorise their recognition as a Remember, Familiar or Guess response, as in

[2] In order to calculate gamma, these ratings were assigned to three classification bins at 25% intervals, i.e. 25% to 49%, 50% to 74%, and 75% to 100%.

Experiment 2.1. Definitions were available in written form throughout the experimental procedure for participants to refer to. As items were presented via Powerpoint, a time limit was imposed for each combined recognition and recollection response of 10 seconds. Piloting ensured this was a reasonable time for all participants to be able to make their recognition and recollection responses without affecting accuracy.

Each group of participants was assigned to one of two versions of the task. Presentation order of items within each stage of Version B was the reverse of that in Version A.

A practice stage was completed prior to commencement of testing. This consisted of learning 10 cue-target pairs, attempting cued recall & FOK, followed by recognition & recollection. This ensured understanding of each stage of the study, and of the FOK likelihood ratings and the recollection definitions. As justifications of RFG responses were not obtained, it was of special importance to ensure the recollection definitions were understood by all participants. The placement of the two hour break was also emphasised during the practice stage.

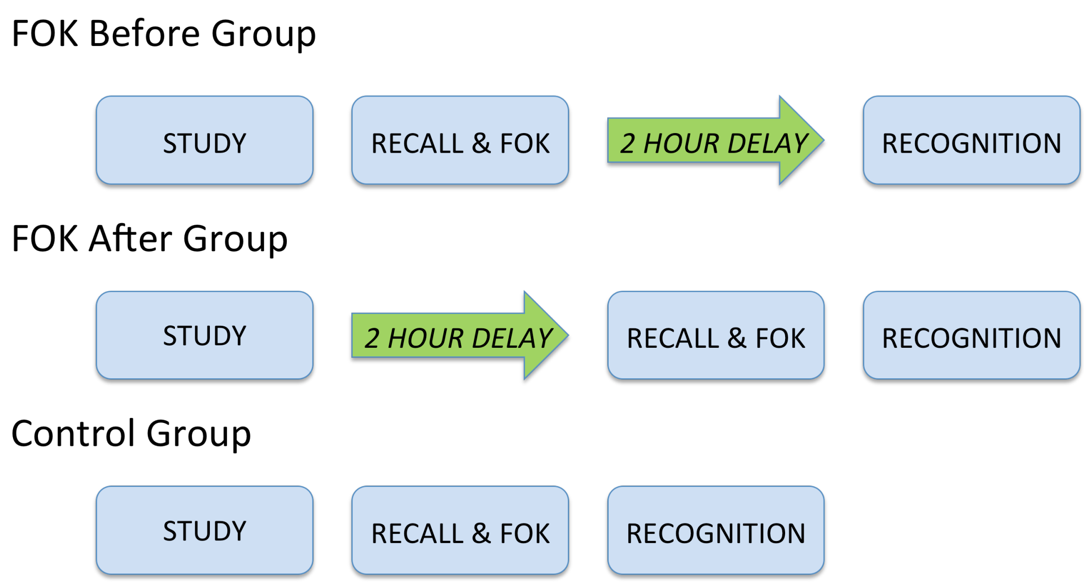


Figure 3.1: Graphic outlining the different stages of the procedure in each condition for Experiment 3.1.

Specific procedures. See Figure 3.1 for a stage by stage diagram of each condition. For the Control condition, having gained informed consent participants completed the practice phase to explain each stage of the experiment. Following practice, participants began the experiment and were presented with the 50 cue-target pairs for learning. Cued recall and FOK ratings were then obtained, and immediately after this the recognition and recollection stage was completed. The FOK Before group also completed the practice phase before commencing the learning stage. Subsequent to learning, the FOK Before group then gave recall and FOK ratings for the cue words. The two hour break was then given, where participants were free to leave the experimental room with instructions not to rehearse the cue-target pairs as the study was assessing forgetting. Upon return, the FOK Before group completed the recognition and recollection stage. The FOK After group completed practice and learning as in the other two conditions. After learning, the two hour break took place, with participants receiving the same instruction not to rehearse during the break. Upon return, cued recall & FOK ratings were obtained, followed immediately by the recognition and recollection measures. All participants were debriefed upon completion of the experimental procedure.

3.2.3 Results

A minimum of 10 items on which FOK judgements are made is desirable for accurate calculation of FOK accuracy. Two participants were excluded due to high levels of recall, leading to fewer than 10 items available for assessment of FOK accuracy. Six further participants were excluded due to inappropriate use of the rating scale, providing a rating of 25 to all unrecalled items. Without variation in FOK rating category use it is not possible to ascertain the accuracy of FOK judgements. The remaining 60 participants (age range 18-30, M = 21.57, SD = 3.54, 15 male) were included in the analyses reported below, with 20 participants assigned to each condition.

3.2.3.1 Memory

A one way ANOVA was used to examine the effect of the two hour delay on recall performance. The proportion of items correctly recalled for all items was calculated by dividing the number of correct responses by the total number of possible responses (60) (see Figure 3.2). A significant effect of group was observed, $F(2, 57) = 9.025$, $p < .001$, $\eta^2 = 0.49$, with Bonferroni post hocs revealing that the FOK After group recalled significantly fewer words than the FOK Before ($p = .009$) and Control groups ($p < .001$). No difference was found between the FOK Before and Control groups. As expected, the delay between learning and recall experienced by the FOK After group led to a significant reduction in recall accuracy compared to the two groups who gave recall responses immediately following learning. Recognition accuracy was examined as the proportion of correct responses out of the total number of possible responses. A significant effect of group was found $F(2, 57) = 5.894$, $p = .005$, $\eta^2 = 0.41$. Bonferroni post hocs revealed that the FOK Before and FOK After groups showed similar levels of recognition performance. In contrast, the Control group was more accurate than either of the experimental conditions ($p = .006$ and $p = .036$ respectively), indicating that the experimental manipulation of delay did affect both recall and recognition measures of memory as expected.

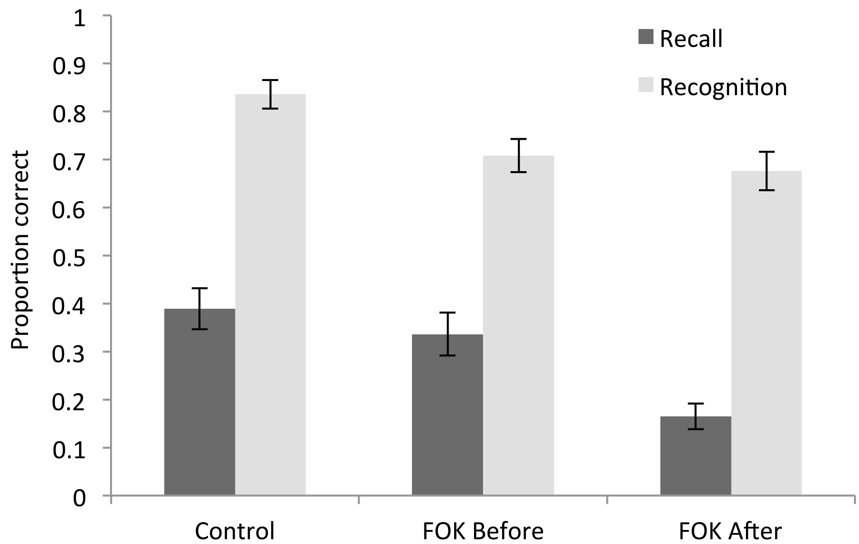


Figure 3.2: Proportion of total items presented which were correctly recalled and recognised for each condition

Omission and commission errors were also examined with one way ANOVAs (see Table 3.1). A significant effect of group was observed for omission errors, $F(2,57) = 9.990$, $p < .000$, $\eta^2 = .260$. Bonferroni posthocs revealed that the FOK After group made significantly more omissions than the Control group ($p < .001$) and the FOK Before group ($p = .002$). No difference was detected between the Control and FOK Before groups. An effect of group was also observed for commission errors, $F(2,57) = 3.815$, $p = .028$, $\eta^2 = .118$. In this instance, the Control group made significantly fewer commission errors than the FOK Before group ($p = .037$), with the FOK After group giving a proportion of commission errors between the two.

Table 3.1: Proportion of omission and commission errors made during the recall test by each participant group

	Control	**FOK Before**	**FOK After**
	M (SD)	***M (SD)***	***M (SD)***
Omission errors	0.574 (0.188)	0.593 (0.200)	0.792 (0.111)
Commission errors	0.037 (0.030)	0.071 (0.049)	0.043 (0.044)

3.2.3.2 Metacognition

The overall magnitude of the FOK judgements for all unrecalled items regardless of subsequent recognition was analysed first (see Table 3.2). An effect of group was found $F(2,57) = 4.809$, $p = .012$, $\eta^2 = 0.38$. Bonferroni post hocs revealed that this effect was due to the FOK Before group showing significantly lower FOK ratings than the Control group ($p = .009$), which could indicate that although they are making FOK judgements at the same stage in the procedure as the Control group, the FOK Before group have reduced their ratings to take into account the upcoming two hour delay. No differences were shown between the FOK Before and the FOK After groups, or between the FOK After and Control groups. Ratings were also grouped into three bins (25% to 49%, 50% to 74%, 75% to 100%) to allow computation of gamma. The distribution of ratings across these bins was examined with a 3 (group) x 3 (rating bin) mixed ANOVA (see Figure 3.3). A main effect of rating category was found, $F(2,114) = 51.619$, $p < .001$, $\eta^2 = .475$, with Bonferroni posthocs revealing that the lowest category of 25% to 49% was used more than the next category of 50% to 74% ($p < .001$), which in turn was used significantly more than the highest category of 75% to 100% ($p = .042$). A marginal interaction was also found between group and rating, $F(4,114) = 2.560$, $p = .061$, $\eta^2 = .082$. One way ANOVAs revealed that the main differences lay between the Control group and the FOK Before group, with the Control group making less use of the 25% to 49% category but more use of the 75% to 100% category than the FOK Before group. Thus the anticipation of the two hour delay in the FOK Before group leads them to show underconfidence in their later recognition performance compared to the Control group who know that no delay will occur before the recognition test. Interestingly, the nonsignificant difference in bin use for the FOK After group suggests that the FOK Before group overcompensate for the upcoming delay and reduce their FOK predictions more than is necessary.

Table 3.2: Magnitude of FOK ratings

	Control *M (SD)*	**FOK Before** *M (SD)*	**FOK After** *M (SD)*
FOK rating *Unrecalled, all items*	50.701 (13.609)	39.431 (9.877)	44.833 (10.661)
FOK rating *Unrecalled, recognised*	52.130 (14.781)	41.148 (11.804)	46.072 (11.257)
FOK rating *Unrecalled, not recognised*	42.438 (10.632)	37.248 (9.669)	41.554 (11.562)

The magnitude of ratings for unrecalled items which were subsequently recognised or not was next considered. As with the Yes/No FOK analysis in Chapter 2, this provides a measure of accuracy; if FOKs are diagnostic of future recognition it would be expected that the magnitude of ratings would be higher for those items which were later recognised than those which were not (Table 3.2). A 3 (group) x 2 (status: recognised or not recognised) mixed ANOVA revealed no main effect of group was present $F(2,55) = 2.342$, $p = .106$, $\eta^2 = .078$. Two participants in the Control group correctly recognised all unrecalled items, thus were not able to be included in the current ANOVA. Thus the main effect of group previously found is removed when considering subsequent recognition. A main effect of status is observed, $F(2,55) = 20.013$, $p < .001$, eta = .267, indicating that correctly recognised items did indeed receive higher ratings than those items which were not recognised. No interaction between group and status was present $F(2,55) = 1.125$, $p = .332$, $\eta^2 = .039$. Based on this measure of accuracy, the placement of the two hour delay did not influence how well participants were able to judge future recognition performance for unrecalled items[3].

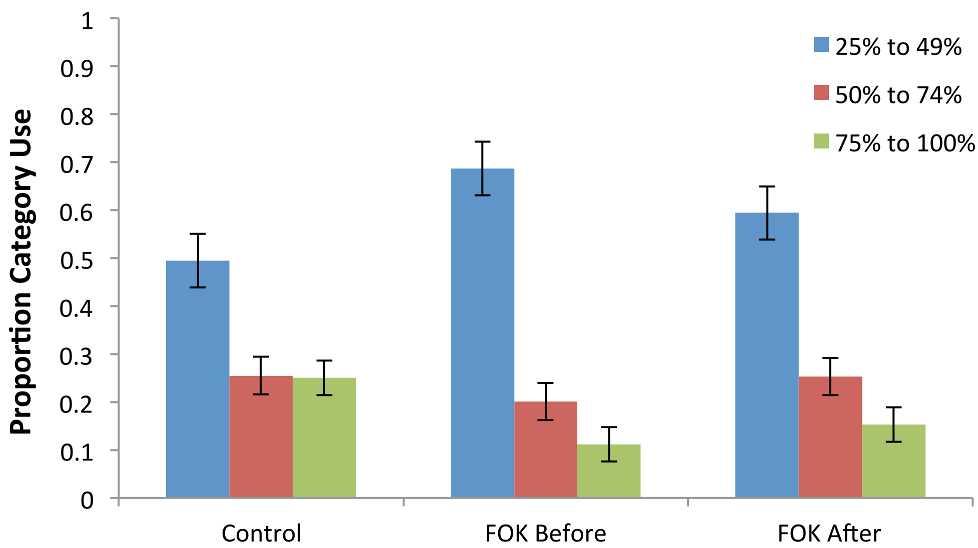


Figure 3.3: Proportions of ratings given for unrecalled items in each bin for each condition

Gamma correlations were computed for unrecalled items only. When examined with a one-way ANOVA, no main effect of group was observed for gamma correlations, $F<1$, indicating that the two hour delay did not affect participants ability to accurately predict their future recognition performance. One sample t-

---

[3] A 3(group) x 3(rating bin) x 2(recognised or not recognised) ANOVA could not be calculated due to missing data for all but 24 participants.

tests against zero showed that the Control group were significantly above chance (*M* = 0.319, *SD* = 0.439, *t*(19) = 2.993, *p* = .009), indicating some level of predictive accuracy for their FOK ratings. However, neither the FOK Before group (*M* = 0.269, *SD* = 0.624, *t*(19) = 1.827, *p* = .085) nor the FOK After group (*M* = 0.221, *SD* = 0.481, *t*(19) = 2.051, *p* = .054) reached significance, although both groups could be considered marginally above chance.

In summary, the effects of the delay on measures of memory accuracy are as expected, with accuracy lower following a delay. No clear effects are present on FOK magnitude or accuracy however. FOK magnitude does seem to be influenced somewhat by the knowledge of an upcoming delay, but is generally quite stable over each condition.

The relationship between FOK accuracy and memory performance was further examined with Pearson bivariate correlations. When all groups were considered together, medium positive correlations between gamma and recall accuracy, *r*(55) = .282, *p* = .037, and gamma and recognition accuracy, *r*(55) = .296, *p* = .028, indicate that as memory improves so does the accuracy of FOK predictions. When each group was examined separately, only the Control group showed a relationship between memory and gamma, with a marginal correlation for recall, *r*(17) = .460, *p* = .063, and a significant correlation for recognition, *r*(17) = .529, *p* = .029. The FOK Before group showed no correlation between gamma and recall, *r*(18) = .204, *p* = .417, or between gamma and recognition, *r*(18) = .236, *p* = .346, nor did the FOK After group, recall *r*(20) = .245, *p* = .297 and recognition *r*(20) = .217, *p* = .358.

3.2.3.3 Recollection

The proportion of R and F responses for hits was examined as in Experiment 2.1 (see Table 3.3). A 3 (group) x 2 (state: R and F) mixed ANOVA revealed no main effect of group *F*(2,57) = 2.175, *p* = .123, η^2 = .071. A main effect of state did occur, *F*(2,57) = 61.014, *p* < .001, η^2 = .517, with more R responses given than F responses. No interaction between group and delay was found, *F*(2,57) = 1.867, *p* = .164, η^2 =

.061. Thus the manipulation of delay did not affect the levels of R and F experienced by the participants.

Table 3.3: Proportions of R and F responses for recognition hits only

	Control *M (SD)*	FOK Before *M (SD)*	FOK After *M (SD)*
Proportion R responses	0.588 (0.210)	0.587 (0.018)	0.486 (0.152)
Proportion F responses	0.279 (0.096)	0.245 (0.112)	0.303 (0.123)
FOK-RF gamma	0.166 (0.275)	0.191 (0.430)	0.282 (0.336)

A gamma correlation was also calculated between RFG and FOK (see Table 3.3). A one way ANOVA revealed no effect of group F<1, indicating that delay placement had no effect on the relationship between FOK and recollection. One sample t-tests demonstrated correlations significantly above chance for the Control group, $t(18) = 2.637$, $p = .017$, and the FOK After group, $t(19) = 3.748$, $p = .001$, and marginally above chance for the FOK Before group, $t(17) = 1.884$, $p = .077$. This possibly indicates that both the FOK and recollection rely on similar experience based judgements. When the FOK judgement is made prior to the delay, recollection is still experience based however the FOK is also influenced by theory based judgements, leading to a slight reduction in the association of the two subjective experiences.

3.2.4 Discussion

The aim of Experiment 3.1 was to examine the effect of a delay on FOK judgements. In contrast to previous research, in the current study the placement of the delay within the standard FOK paradigm was manipulated in order to alter the reliance of FOK on theory and experience based judgements, similar to that previously shown with JOLs (Koriat et al., 2004). The very nature of the FOK, being a judgement about future memory performance for an unrecalled item, means that online assessment of partial information will always form a basis of the judgements However, the placement of the delay prior to the recognition test necessitates the recruitment of theory based information as well to increase the accuracy of the FOK judgement. Recall and recognition indicated that the two hour delay had the predicted effect on memory, with accuracy of each measure decreasing following the delay. Magnitude of FOK judgements also showed an effect of delay, with the FOK Before group predicting much lower FOK ratings than the control group. However, no effect of delay was observed for the accuracy of FOK judgements, whether measured by comparing the magnitude of FOK responses for correctly and incorrectly recognised items, or as measured by the gamma correlation. This shows a clear dissociation between memory performance and FOK accuracy.

The pattern of results shown for the FOK magnitude could show the recruitment of theory based judgements in addition to information based judgements when a delay is introduced between FOK and recognition. The FOK Before group reduce their FOK predictions relative to the Control. As both groups have experienced the same procedure and show the same levels of recall accuracy, online assessment of partial information for unrecalled items will also be similar. The difference in magnitude is therefore a result of theory based judgements, the anticipation of the effects of the delay for the FOK Before group. Interestingly, the FOK magnitude for the FOK After group lies between these two. This group has already experienced the two hour delay and can make a direct assessment of their current memory ability to establish how much decline has occurred, thus do not need to incorporate theory based judgements. It would therefore appear that, although the FOK Before group are correct to reduce their FOK ratings compared to the Control group, they overestimate the effects of the delay, resulting in the significant difference between the Control and FOK Before groups with no effect for the FOK After group. On the surface, these findings may be viewed as being somewhat at odds with the JOL literature (Koriat et al., 2004), where an effect of delay and the influence of theory based judgements were only detectable for comparative judgements. The single judgement obtained here should therefore show no effect of delay. The key difference between JOLs and FOKs is the stage at which they are made. JOLs are made prior to recall, thus no explicit retrieval attempt is made and subsequently the partial information activated during retrieval will not be available. The FOK in contrast is made following a failed retrieval attempt, and therefore will always be influenced by the partial cues accessed during the failed recall attempt. The placement of the delay can vary the reliance on this information, and how much theory based judgements are recruited in addition.

With regards to FOK accuracy, the results support the predictions of the MCH (Hertzog et al., 2010). All participants experience the same encoding conditions, so despite the implementation of a delay there should be no effect on FOK accuracy. The assumption of equivalent encoding is supported by the levels of recall and recognition accuracy. When tested at the same point no effects of group are shown, with recall accuracy showing no group differences for the Control and FOK Before groups, and recognition accuracy showing no group differences for the FOK Before and FOK After groups. With equivalent encoding between the three groups, all groups have access to similarly diagnostic information on which to base their FOK judgement, leading to similar levels of accuracy.

Closer examination of the gamma correlations does indicate a slight effect of delay. When comparing to chance performance, only the Control group were found to have above chance FOK predictive accuracy, with the FOK After group being marginally above chance, and the FOK Before group even less marginally above chance. This pattern of results could be explained by the accessibility account (Koriat, 1993). The FOK After group have experienced the two hour delay, thus their FOK judgements are based on the quality and quantity of partial information available following a period of forgetting. This will presumably be lower than that available to the Control group, as forgetting will not only affect retrievability of the target but also retrievability of the partial information. While the accessibility account would predict this would affect the magnitude of FOKs through less information being available, as found in the current study, it may also affect FOK accuracy. Reduced availability of partial information may affect the balance of correct and incorrect partial information that is accessed. Thus if incorrect information is more susceptible to forgetting, the FOK accuracy may increase as the accessible details are more informative to future recognition. Alternatively, more correct information may be forgotten, reducing FOK accuracy, or it may not affect the balance of correct and incorrect information available. This could be ascertained by replicating the current study with an included assessment of partial information available for report. A second explanation could lie in the observation that the pattern of FOK accuracy counters that of errors of commission, thus changes in task difficulty may be influencing predictive ability. The Control group make the fewest commission errors while the FOK Before group make the highest commission errors. More errors imply a lower criterion for recall, thus items with lower memory strength are reported and so are no longer available for FOK predictions. The remaining items will thus be closer together in terms of memory strength, making it harder to discriminate between those which will be subsequently recognised and those which will not. Differences in FOK accuracy may therefore be a consequence of differences in task difficulty at the FOK stage.

The present study also examined the effect of delay on recollection and its relationship with FOK performance. It was unknown whether the intermediate delay of two hours would affect the level of Remember responses given, as previous research has focused on either very short delays of a few minutes which have no impact on Remember responses (Yonelinas & Levy, 2002) or much longer delays measured in days which lead to a reduction in Remember responses (Knowlton & Squire, 1995). No effect of delay was observed within the current paradigm, indicating that recollection is not reduced over a delay of only two hours. Interestingly, Familiar responses were also unaffected by the delay. This is surprising, as previous research has found that familiarity is particularly susceptible to forgetting and levels begin to decline after very short delays (Hockley, 1992). The associative recognition task used here may explain this finding, as memory for new associations has been shown to be supported by recollection and not familiarity (Hockley, 1992; Yonelinas, 1999). The contribution of familiarity in the present study is therefore limited by the design of the study, thus it may not be possible to detect any changes in the level of Familiar responses due to forgetting. In addition, recollection and delay have previously been investigated using within subjects designs. The current study is between subjects. The type of design used can have significant impact on the results obtained, as discussed in Chapter 2. Of greater interest is the relationship between FOK and recollection. Gamma correlations establish a small but significant relationship between FOK and recollection, such that increased FOKs were associated with increased recollection. This is in agreement with the findings of Hicks & Marsh (2002), although the correlations reported here are somewhat lower than their finding of an overall gamma of 0.35. Nonetheless, the presence of a relationship between the two subjective states adds support to the proposal of Souchay et al. (2007) that FOK and recollection rely on similar partial information that is retrieved during recall and recognition attempts.

3.3 Experiment 3.2: Delay and FOK in ageing – within subjects

3.3.1 Introduction

The recruitment of theory based information to inform the FOK was demonstrated by placing a delay between the FOK and recognition performance. This indicates that people overestimate the impact of forgetting over a two hour time interval, shown by lower FOK judgements when made prior to the delay than after the delay. The present study aims to replicate this finding using a within subjects design. The experimental design used can have unexpected effects on the pattern of results shown in metacognitive studies. For example, research into the effect of overlearning and FOK demonstrates no effect of repeated encoding events on FOK if a between subjects design is used (Carroll & Simington, 1986), whereas within subjects designs demonstrate an increase in FOK accuracy as learning increases (Nelson et al., 1982). This is due to the increased variability inherent in between subjects manipulations. Each participant will have a slightly different criterion placement for deciding whether an item will be retrievable in future or not, resulting in variability which may mask the effects of the intended manipulation. By using a within subjects design, this variability is reduced, making the experimental manipulation easier to detect (Carroll & Nelson, 1993). It is therefore possible that the effects demonstrated in Experiment 3.1 underestimate the influence of theory based judgements on the FOK. The aim of the current study was to establish if this is indeed the case.

In addition to replicating the effect of delay on FOK, the influence of ageing was also examined. As previously discussed, older adults demonstrate impaired memory performance compared to young adults (Anderson & Craik, 2000). One contributor to this impaired memory performance is the inability of older adults to spontaneously use appropriate strategies during encoding to facilitate memory (for an overview see Lemaire, 2010). For example, Souchay & Isingrini (2004) demonstrated that when no time limit was imposed during encoding, young adults took advantage of this to increase their study time compared to an experimenter paced task, and subsequently improve their recall performance. Older adults, in contrast, did not increase their study time and so showed no increase in recall performance. However, when prompted to utilise the appropriate memory strategies by the experimenter, older adults are able to effectively apply and use these strategies to improve their memory accuracy (Froger, Bouazzaoui, Isingrini, & Taconnat, 2012; Murphy et al., 1981, 1987). Similarly, Thomas et al. (2011) demonstrated that older participants did not spontaneously use the partial information retrieved in a failed recall attempt to inform their FOK judgements. Only by explicitly prompting older adults to report accessible partial information prior to making an FOK were they then able to use this information as a basis for their judgements. Young adults were able to do this with no prompting required. This demonstrated failure of older adults to automatically use their available knowledge of strategies and the available diagnostic information to improve their memory ability raises the possibility that they may also fail to automatically account for the effects of a delay on memory accuracy. The present study will examine the pattern of FOK ratings for young and older adults rather than manipulate instructions. If the same pattern is shown between the two age groups, and mirrors the pattern of results in Experiment 3.1, then we may assume that both age groups have spontaneously employed their beliefs about the effects of delay on memory to adjust their FOK ratings accordingly. If, however, an age difference appears, it may suggest that older adults are not able to spontaneously employ this information to adjust their FOK judgements.

3.3.2 Method

3.3.2.1 Participants

The young adult group consisted of 21 undergraduates from the University of Leeds (age range 18 to 25, M = 21.24, SD = 3.25) who participated in return for course credit. Young adults had not previously taken part in Experiment 3.1. Twenty four older adult volunteers were recruited from the local community (age range 62 to 84, M = 71.27, SD = 6.66). All older adults in the study attained scores on the Mini Mental State Examination above the cut off of 27 (M = 29.21, SD = 1.06) and were not taking medication that would affect cognitive function. All participants received more than nine years of education. However, a marginal difference was observed between the two age groups (young adults M = 15.67, SD = 2.08; older adults M= 14.33, SD = 2.90; $t(43) = 1.749$, p = .087), with older adults reporting slightly fewer years of education than the young adults.

3.3.2.2 Materials and Procedure

Materials were the same as those used for Experiment 3.1, consisting of 50 unrelated word pairs matched on frequency and concreteness. The general procedure was also the same, with the addition of a second recall & FOK stage (see Figure 3.4). Participants began the study with the learning phase, with the 50 word pairs presented for 5 seconds each via Powerpoint, using Calibri font size 60. Following this, the first cued recall & FOK stage was completed, immediately after learning. The two hour break was then implemented, with the instruction not to rehearse the word pairs. Upon return, a second cued recall & FOK phase was completed by the participants. Finally, the recognition & recollection phase was carried out as previously, with the cue word presented in lower case and 4 response option presented in upper case (Calibri font size 48).

Figure 3.4: Graphic outlining the stages of the procedure for Experiment 3.2

3.3.3 Results

3.3.3.1 Memory

Analysis and calculations of proportions follows that of Experiment 3.1. First, the proportion of items correctly recalled was examined. A 2 (age: young adults and older adults) x 2 (time: immediate and delay) mixed ANOVA was used (see Figure 3.5). A main effect of age was present, $F(1,43) = 4.581$, $p = .038$, $\eta^2 = 0.10$, with older adults recalling fewer items than young adults. A main effect of time was also found, with recall significantly lower after the two hour break than before, $F(1,43) = 31.871$, $p < .001$, $\eta^2 = 0.43$. No interaction was observed, $F<1$, indicating that the delay did not lead to a disproportionate effect on recall for either age group. Recognition performance was compared with an independent samples t-test. A marginal effect of age was observed $t(43) = 1.802$, $p = .079$, $\eta^2 = 0.26$, with older adults showing slightly lower accuracy ($M = .680$, $SD = .186$) than young adults ($M = .774$, $SD = .162$).

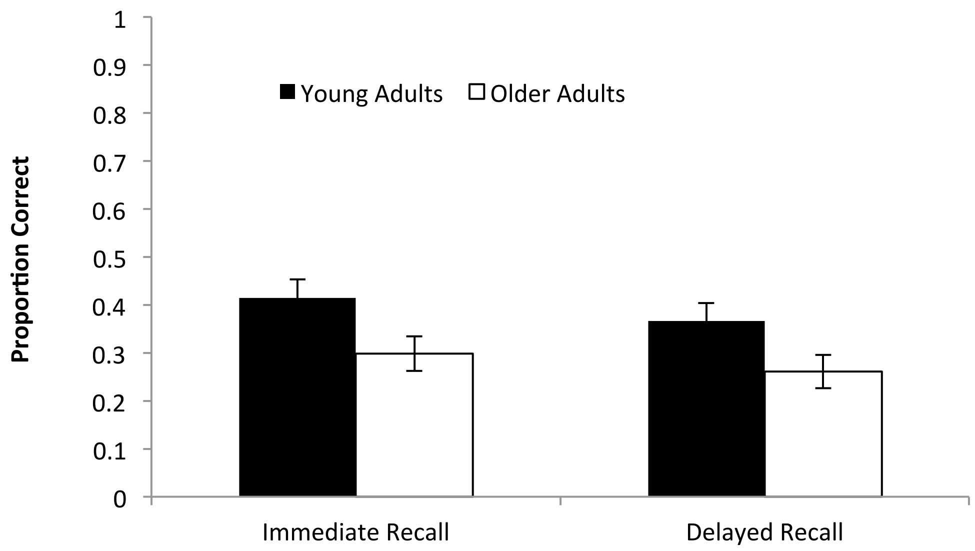


Figure 3.5: Proportion of items correctly recalled by young and older adults when tested immediately after learning and after a two hour delay

Omission errors were examined with a 2 (age) x 2 (time) mixed ANOVA (see Table 3.4). No main effect of age was found, $F(1,43) = 2.138$, $p = .151$, $\eta^2 = .047$. A main effect of time was present, $F(1,43) = 38.823$, $p < .001$, $\eta^2 = .461$, with more omission errors occurring following the 2 hour delay than prior to the delay. No interaction between age and time was detected, $F(1,43) = 1.289$, $p = .263$, $\eta^2 = .029$. Commission errors were also subject to a 2 (Age) x 2 (time) ANOVA (see Table 3.4). Here, a main effect of age did occur, $F(1,43) = 4.963$, $p = .036$, $\eta^2 = .098$, with older adults making more commission errors than young adults. Again, a main effect of time was also present, $F(1,43) = 4.284$, $p = .045$, $\eta^2 = .091$. Unlike omission errors, participants were less likely to give commission errors following the 2 hour delay. No interaction between age and time was present, $F<1$.

Table 3.4: Proportion of omission and commission errors made by young and older adults at each time point

		Young adults M (SD)	Older adults M (SD)
Time 1: Immediate	Omission errors	0.536 (0.186)	0.628 (0.198)
	Commission errors	0.050 (0.043)	0.074 (0.054)
Time 2: Delayed	Omission errors	0.602 (0.172)	0.673 (0.199)
	Commission errors	0.032 (0.040)	0.066 (0.057)

3.3.3.2 Metacognition

The magnitude of FOK judgements was examined using a 2 (age) x 2 (time) mixed ANOVA (see Table 3.5). No effect of age was found, F<1, however a main effect of time was observed, $F(1,43) = 5.447$, $p = .024$, $\eta^2 = 0.11$. Although the means appear quite similar, FOK ratings were significantly lower after the two hour delay. No interaction between age and time was present, F<1. The distribution of ratings across the three bins was also examined (see Figure 3.6), as in Experiment 3.1, with a 2 (age) x 2 (time) x 3 (rating bin) mixed ANOVA. A main effect of rating bin was found, $F(2,86) = 71.228$, $p < .001$, $\eta^2 = .624$, with the lowest rating category used to the greatest extent and the highest rating category used the least. No interactions were present (age and rating bin: F<1; time and rating bin: $F(2,86) = 2.624$, $p = .097$, $\eta^2 = .058$; age, time and rating bin: $F(2,86) = 1.166$, $p = .304$, $\eta^2 = .026$).

Thus far, the results have found the expected effects of age and delay on memory performance. In addition, there is an effect of delay but not of age on FOK magnitude. FOK accuracy will be considered next.

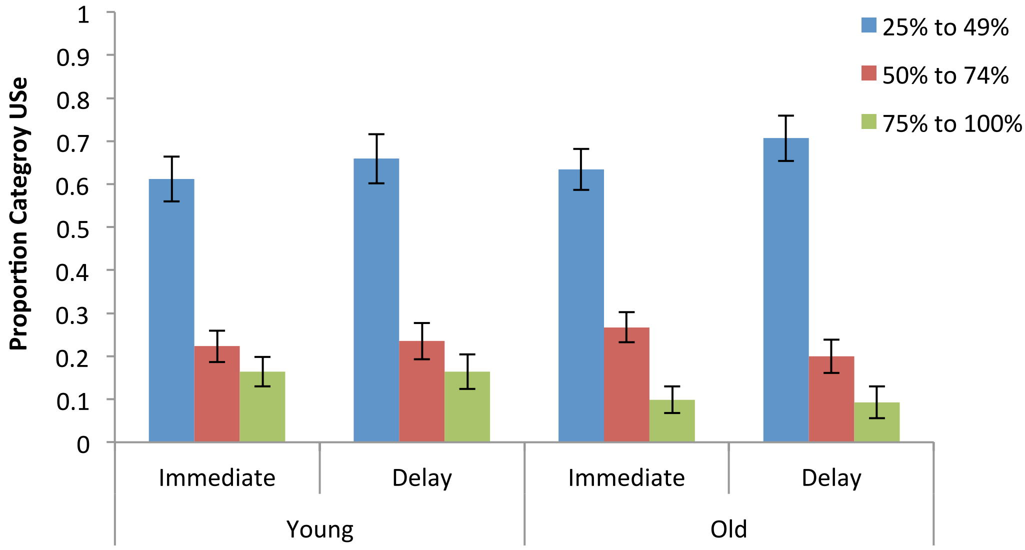


Figure 3.6: Proportion category use of FOK ratings for unrecalled items for young and older adults

The accuracy of FOK ratings was first examined by comparing the ratings for items which were subsequently recognised to those which were not recognised (Table 3.5). A 2 (age) x 2 (time) x 2 (recognition accuracy: correct and incorrect) ANOVA revealed no effect of age group $F<1$, or of time $F(1,43) = 2.357$, $p = .132$, $\eta^2 = .052$. A main effect of recognition accuracy was found, $F(1,43) = 14.451$, $p < .001$, $\eta^2 = .252$, with items which were later recognised receiving higher FOK ratings than those items which were not recognised. Thus FOK ratings are predictive of future recognition by this analysis. No interaction reached significance (age and time: $F(1,43) = 2.145$, $p = .150$, $\eta^2 = .048$; age and recognition accuracy: $F<1$, time and recognition accuracy: $F<1$; age, time and recognition accuracy $F(1,43) = 1.650$, $p = .206$, $\eta^2 = .037$).

Table 3.5: Magnitude of FOK Ratings for Young and Older Adults

		Young adults *M (SD)*	Older adults *M (SD)*
Time 1: Immediate	FOK rating ***Unrecalled, all***	42.648 (11.018)	40.501 (9.124)
	FOK rating ***Unrecalled, recognised***	43.434 (11.274)	41.255 (9.365)
	FOK rating ***Unrecalled, not recognised***	39.585 (13.091)	38.380 (11.811)
Time 2: Delayed	FOK rating ***Unrecalled, all***	41.258 (12.914)	37.774 (10.627)
	FOK rating ***Unrecalled, recognised***	42.134 (12.403)	38.809 (9.651)
	FOK rating ***Unrecalled, not recognised***	40.748 (15.329)	34.954 (9.863)

Gamma correlations were also considered (see Table 3.6). No effect of age $F(1,43) = 1.763$, $p = .191$, $\eta^2 = .039$, or time $F<1$ was present. An interaction between age and time was observed $F(1,43) = 6.219$, $p = .017$, $\eta^2 = .126$. Follow up t-tests revealed that the accuracy of FOK judgements made immediately after learning was equivalent for the two age groups $t(43) = 0.544$, $p = .589$, whereas after the two hour delay a significant difference emerges $t(43) = 2.520$, $p = .016$, with the older adults achieving a higher gamma correlation than the young adults. Indeed, while the young adults stay consistent with their ratings showing no difference between immediate and delayed FOK accuracy, $t(20) = 1.167$, $p = .257$, older adults show higher accuracy after the two hour delay than before the two hour delay $t(23) = 2.943$, $p = .007$.

One sample t-test against zero revealed that older adults' gamma correlations were significantly above chance both prior to $t(23) = 2.545$, $p = .018$, and after $t(23) = 5.587$, $p < .001$ the two hour delay. For the young adults, although the gamma correlation was significantly above chance when made prior to the two hour delay $t(20) = 2.702$, $p = .014$, after the two hour delay gamma fails to reach significance, $t(20) = 0.850$, $p = .405$. This pattern is somewhat surprising, as it would be expected that the young adults would be reliably above chance whereas older adults may or may not be above chance. However, in this instance it would appear that the young adults are adversely affected by the delay to a small extent, whereas older adults are able to maintain accuracy in their FOK ratings.

Table 3.6: Gamma correlations for young and older adults prior to and after the delay

	Young adults ***M (SD)***	**Older adults** ***M (SD)***
Immediate Gamma	0.337 (0.571)	0.251 (0.483)
Delayed Gamma	0.119 (0.644)	0.541 (0.474)

The relationship between memory measures and gamma was next examined. When collapsed across age group, a moderate significant correlation was found between recall performance and immediate gamma accuracy, $r(45) = .302$, $p = .044$. The accuracy of FOKs made prior to the delay increases as recall accuracy increase. No relationship was shown between gamma accuracy after the delay and recall, $r(45) = -.070$, $p = .647$. Likewise, no correlations were found with recognition performance and gamma accuracy, either when the FOK was made immediately, $r(45) = .163$, $p = .285$, or when the FOKs were made post-delay, $r(45) = -.109$, $p = .475$. When each age group is considered separately, however, no correlations reach significance (see Table 3.7).

It was also of interest to examine whether a relationship existed between the magnitude of the FOK given and the level of recall accuracy both before and after the delay. For young adults, a moderate positive correlation was observed between FOK magnitude and recall accuracy prior to the delay, $r(21) = .434$, $p = .049$, and after the delay, $r(21) = .482$, $p = .027$. However, no relationship was found between FOK magnitude at either time point and recognition accuracy, immediate $r(21) = .307$, $p = .176$, delayed $r(21) = .384$, $p = .085$. For older adults, no correlations were observed between FOK and recall either immediately, $r(24) = -.089$, $p = .680$, or after the delay, $r(24) = .121$, $p = .574$. Likewise, FOK did not correlate with recognition accuracy, immediate $r(24) = .059$, $p = .784$, delayed $r(24) = .080$, $p = .709$. The difference in the patterns of correlations found in young and older adults could suggest that these two age groups use different types of information on which to make their FOK predictions. Thus while the effect of delay leads to a reduction in recall accuracy and subsequently, to some extent, FOK accuracy within the young adults, the delay only affects recall accuracy within the older adults as they base their FOKs on different information.

Table 3.7: Correlations between measures of memory accuracy and gamma for young and older adults

		Young adults	**Older adults**
		r (p)	***r (p)***
Recall accuracy	Immediate Gamma	.318	.270
		p = .160	*p = .210*
	Delayed Gamma	-.135	.255
		p = .560	*p = .230*
Recognition accuracy	Immediate Gamma	.243	.063
		p = .288	*p = .770*
	Delayed Gamma	-.150	.122
		p = .516	*p = .571*

N = 21 for young adults, N = 24 for older adults

3.3.3.3 Recollection

The level of R and F responding for hits only was analysed with a 2 (age) x 2 (state: R and F) mixed ANOVA (Figure 3.7). A main effect of age was found, $F(1,43) = 10.217$, $p = .003$, $\eta^2 = 192$, with young adults reporting an experiential state for more trials than older adults. A main effect of state was also observed $F(1,43) = 91.150$, $p < .001$, $\eta^2 = .679$. As previously in Experiment 3.1, more R responses were given then F responses. No interaction was present $F(1,43) = 0.984$, $p = .327$, $\eta^2 = .022$, suggesting that the older adults do not show a selective impairment in R responding only. The main effect of state therefore reflects lower levels of F responding in the older adult group as well as the expected reduction in R responding.

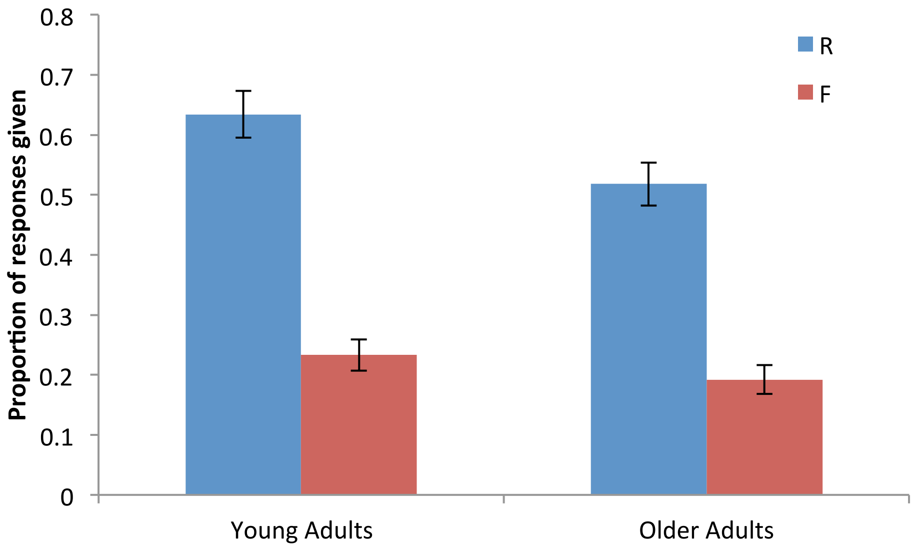


Figure 3.7: Proportion R and F responses given for correctly recognised items only by young and older adults

Gamma correlations between the FOK at each time point and RFG responding were also examined with a 2 (age) x 2 (time) ANOVA. No effect of age was present $F(1, 42) = 0.052$, $p = .821$, $\eta^2 = .001$. A main effect of time was found $F(1,42) = 4.269$, $p = .045$, $\eta^2 = .092$, with higher correlations found after the delay than at immediate test. No interaction was present $F(1,42) = 1.023$, $p = .318$, $\eta^2 = .024$. One sample t-tests replicated these findings, with immediate gammas not significantly different from chance for either age group, young adults $M = 0.139$, $SD = 0.455$, $t(20) =$

1.399, p = .177 and older adults M = 0.113, SD = 0.520, $t(22)$ = 1.040, p = .310. After the delay, young adults' are marginally above chance, M = 0.200, SD = 0.499, $t(20)$ = 1.840, p = .081, whereas older adults are significantly above chance, M = 0.292, SD = 0.574, $t(23)$ = 2.284, p = .032.

3.3.4 Discussion

The aim of the current study was to replicate the effect of delay on FOK judgements and to examine the effect of age within the delay paradigm. The memory results show the expected age effect on accuracy, with older adults consistently showing lower accuracy than the young adults in tests of recall and recognition. In addition, the expected effect of delay occurred, as found in Experiment 3.1. Importantly, no interaction was observed between age and delay, thus older adults were not disproportionately affected by the two hour delay in testing.

Regarding the magnitude of FOK ratings, only an effect of delay was found, with ratings lower following the two hour delay than before the delay. This is the opposite to when tested between subjects, where ratings following the delay were higher than those before. This could reflect differences in application of the belief regarding forgetting and delay. In the within subjects paradigm, at the immediate recall and FOK participants will be trying to predict the amount of forgetting that will occur, as in the FOK Before group from Experiment 3.1. Following the delay, for the between subjects design, the FOK After group do not need to predict the effects of forgetting as they can directly assess it. For the within subjects design of the current study, participants will be conscious that they have already made one judgement, and that a two hour delay has occurred since then, so may therefore further reduce their ratings rather than rely on the available partial information only. By applying their belief about forgetting and delay twice, this then results in lower FOK ratings after the delay. As no effect of age is observed, it can be assumed that older adults are able to use their memory beliefs in a similar way to young adults to adjust their FOK ratings accordingly.

Age effects on FOK accuracy in the present study are somewhat unusual. Although no main effect of age was found, an interaction between age and delay was present, with older adults showing greater FOK accuracy following the delay. Indeed, older adults after the delay were significantly more accurate than young adults. This goes against prediction of both the inferential hypothesis and the MCH. The inferential hypothesis would propose older adults would be less accurate, due to an inability to spontaneously use partial information to inform the FOK judgement (Lemaire, 2010; Thomas, Bulevich, & Dubois, 2011). The MCH would propose that, as older adults show reduced recall accuracy, the FOK judgements would also show reduced predictive accuracy. Furthermore, older adults are ,more likely to produce commission errors, thus reducing the range of memory strength of items available for FOK. Somehow, the 2 hour delay enables older audlts to better discriminate between items which will be recognised and those which will not. Further examination of the effects of delay on FOK in ageing are required to, first, establish whether this effect is replicated and, second, to discover the mechanism behind the improved accuracy of older adults.

Analysis of the recollection data does not shed further light on why no age effect is found for FOK accuracy. As described in Chapter 2, older adults typically show a deficit in recollection, reporting fewer Remember responses than young adults (Bugaiska et al., 2007; Bunce, 2003; Perfect & Dasgupta, 1997; Perfect, Williams, & Anderton-Brown, 1995) and showing deficits in source memory tasks (Dennis et al., 2008; Henkel et al., 1998). It has been suggested that recollection processes rely on the same partial information as the FOK, thus levels of recollection and FOK accuracy are related (Souchay et al., 2007). When examining levels of Remember and Familiar responding in the present study, the key interaction between age and subjective state was not significant. However, older adults did report lower levels of both subjective experiences than young adults. This would suggest that insufficient information is available to elicit feelings of recollection, and therefore there should also be insufficient information for an accurate FOK.

In addition to comparing levels of recollection and FOK, a more direct measure of the relationship between these two subjective states was obtained with gamma correlations. Immediate FOKs are not related to recollection measures, but FOK judgements made after the delay (and therefore immediately before recollective states were reported) showed some association with recollection, especially in older adults. It is to be expected that the partial information and mnemonic cues available at each failed recall attempt are not identical. If recollection experience relies on the same information as the FOK, then it would make sense that the most recently activated information would form the basis of recollective experience, thus the greater relationship between delayed FOK and Remember and Familiar responses displayed here. It is somewhat surprising that gamma correlations for immediate FOKs and recollection are not above chance, as this would imply that the information available at the first failed recall attempt is very different to that which was accessed during the second failed recall attempt. However, the correlations between delayed FOK and recollection remain quite low, suggesting that although some aspect of the two judgements is shared there is a great deal of variation as well, revealing a more complex relationship between the two experiences. As with memory performance, manipulations of recollective experience may not necessarily have the predicted effects on FOK judgements.

In sum, the present study replicated the effect of delay on the magnitude of FOK ratings, with participants reducing their FOK ratings following the two hour delay. Despite clear age effects on memory and recollection, no age effects were shown for FOK accuracy. In fact, older adults were found to have higher FOK accuracy following the delay than young adults. This finding is problematic for both sides of the ageing and FOK debate, as even the MCH would predict age associated reduction in accuracy in light of observed memory deficits. The key difference between the current paradigm and previous studies is the presence of the testing effect, whereby the first recall attempt enables further encoding and improves memory performance. However, with recent research indicating participants are unaware of the benefits of testing on memory (Tullis, Finley, & Benjamin, 2013) it would seem unlikely that this would influence participants' judgements.

3.4 Experiment 3.3: Ageing and environmental support at retrieval I

3.4.1 Introduction

The proposed close link between memory and FOK allows easy exploration of known memory effects within the FOK paradigm to further theory as to how FOKs are formed. Despite thorough consideration of manipulations at encoding, retention interval and retrieval on memory performance, research within the FOK framework has typically focused on encoding (e.g. Carroll & Nelson, 1993; Sacher, Taconnat, Souchay, & Isingrini, 2009), with a few studies considering retention interval as in Experiments 3.1 and 3.2 in the present chapter (Janowsky et al., 1989; Schacter, 1983). Manipulations at retrieval provide an additional interesting avenue to explore the FOK, as differences in the ease of recall of items will presumably also influence the ease with which partial information can also be retrieved, thereby influencing the magnitude and accuracy of the FOK (Koriat, 1995, 1997). The present studies aim to consider how the amount of support provided at retrieval, prior to the FOK, influence the subsequent judgement in young and older adults.

One theory proposed to explain the memory deficit in ageing is the environmental support hypothesis (Craik, 1983, 1986). During learning, the surrounding environment and the context of the memory task itself contribute to the cognitive processes recruited and modify the encoding of the stimuli. Thus memory performance is a result of the interaction between processes initiated by the external environment and those processes initiated by the individual. Ageing leads to greater reliance on external influences due to diminished internal processing resources being available. By maximising external influences, it is then possible to reduce or eliminate the effect of ageing on memory performance. Indeed, studies have shown that when support is provided at both encoding and retrieval the differences in accuracy between young and older adults are significantly reduced (Light, 1991; Naveh-Benjamin et al., 2002).

Environmental support at encoding has been examined within the FOK paradigm, from levels of processing manipulations (e.g. Lupker, Harbluk, & Patrick, 1991) and repeated presentation of items (e.g. Carroll & Nelson, 1993), to name a few. Environmental support at retrieval has not been thoroughly examined however. Within ageing specifically, Thomas et al. (2011) demonstrated that when prompting older participants to retrieve partial information prior to making an FOK, their FOK judgements were greater in accuracy compared to when no prompt was given. The external support given by explicitly requiring older adults to access and report the partial information available led to increased metacognitive accuracy.

The quantity and quality of partial information available after a failed recall attempt determines the magnitude and accuracy of the FOK, as assessed by asking participants to report extraneous details, such as the valence of the target word. In this novel design, the aim was to examine how the provision of partial information affects FOK judgements by manipulating the number of cues given to elicit the target item. According to the environmental support hypothesis, the greater the number of cues the greater the memory performance, which should subsequently lead to better quality and quantity of partial information available on which to base the FOK of unrecalled items. Thus greater environmental support at retrieval should lead to higher FOK magnitude. In addition, if the quality of partial information is improved, this may lead to higher levels of FOK accuracy due to increased diagnostic value. By having one condition with fewer cues provided, it is also possible to examine the effect of spontaneously retrieved partial information on the FOK for these items. Thus participants were presented with five items which they were asked to encode as a single entity. These groupings of items are subsequently referred to as arrays. Each array consisted of a number, a five letter word, a name, a symbol, and a fruit or vegetable. The number of cues for each array could therefore be easily manipulated.

3.4.2 Method

3.4.2.1 Participants

Participants were 41 A-level and undergraduate Psychology students (age range 16-18, *M* = 17.02, *SD* = 0.72) and 20 older adults (age range 61 to 81, *M* = 71.45, *SD* = 6.30). A-level students were recruited as part of a Research Open Day, and, in addition to individual consent at the time of participation, parental consent was obtained via the students' school prior to the research open day. Undergraduate students were recruited from the participant pool maintained by the University of Leeds and took part in return for course credit. Older adults were recruited from volunteers in the local community and were screened for dementia with the MMSE. All older adults achieved above the cut off score of 27 (*M* = 28.90, *SD* = 0.85). Ethical approval for the study was obtained from the Institute of Psychological Sciences Ethics Committee, University of Leeds.

3.4.2.2 Materials

A total of 30 arrays were compiled, with each array consisting of a two digit number, the name of a fruit or vegetable, a five letter word, a male or female name, and a symbol (see Figure 3.8). These categories were decided upon to allow examination of a variety of information, and to minimise participants' use of mnemonic strategies. Two digit numbers were obtained from a random number generator. Fruits and vegetables were identified by asking three pilot participants, who did not take part in the study, to list 20 fruits and 20 vegetables each. The top 15 fruits and top 15 vegetables were then used as stimuli. Names were taken from a list of the most popular male and female names in 1940 and 2010. Words were obtained from the MRC Psycholinguistics Database. Words were limited to five letters and were of high frequency (*M* = 227.5, *SD* = 159.6, Kucera Francis written frequency), high imageability (*M* = 548.4, *SD* = 67.3) and high concreteness (*M* = 528.3, *SD* = 68.8). Finally, symbols were randomly selected from an online database (Symbols.com). An equal number of each type of stimuli were selected as a recall target (i.e. six numbers, six fruit or vegetables, six words, six names and six symbols). A further 18 exemplars for each category were also obtained to provide novel distracters for the recognition task. Piloting ascertained that presenting 15 arrays per learning trial were of sufficient difficulty to ensure recall could occur and that recognition and metacognitive judgements were not at floor level. Therefore the 30 arrays were randomly assigned to one of two blocks: list A and list B.

3.4.2.3 Procedure

Participants were tested in groups of between two and 15 individuals. Answer booklets with written instructions for the study were provided, and these instructions were reiterated by the experimenter prior to commencement of the study. The instructions gave information on each phase of the study and also provided examples of how the recall, FOK and recognition tasks would be presented. Time was allowed for questions before the learning phase began, and participants were asked not to move forward in the answer booklets until instructed to do so. During learning, arrays were presented using PowerPoint 2007 via projector for 12 seconds each, with a one second fixation appearing between each array. Items were presented in the same location for each array, with numbers presented in the top left of the slide, the fruit/vegetable presented top right, the five letter word presented in the centre, the name presented in the bottom left and the symbol was presented in the bottom right (see Figure 3.8). Calibri font size 66 was used for numbers, fruit/vegetables, names and five letter words. Symbol images were scaled to cover an area no larger than 5cm by 5cm.

Following presentation of the first list of 15 arrays, the recall test took place. Each array was printed in the answer booklet as it appeared during the learning phase, with the target item indicated by an arrow. The appropriate number of supporting cues was also given (see Figure 3.8), with remaining placeholders left blank. Participants were asked to recall as much of the missing information from the array as possible, including both target and non-target items. Beneath each array a feeling of knowing scale was also presented, with the prompt 'How confident are you that you will be able to recognise the missing target above if shown the array and target together?'. Confidence levels were listed at 0%, 20%, 40%, 60%, 80% and 100%, with participants asked to circle the appropriate number for all arrays, whether they recalled the target or not. Once recall and FOK responses had been obtained for all arrays, the four alternative forced choice recognition test was then given. Identical arrays to those used in the recall task were presented in the answer booklet, with four possible targets listed below the array. Targets were randomly assigned as to whether they were presented first in the list, second, third or fourth. Participants were asked to circle the appropriate target from the options provided. Following completion of the recognition task, the second set of arrays was presented for learning via Powerpoint, with recall and recognition tested via the answer booklet as above. No time limit was imposed for the recall and recognition tasks. The experimenter waited for all participants to complete each task before allowing the group to continue to the next part of the experiment.

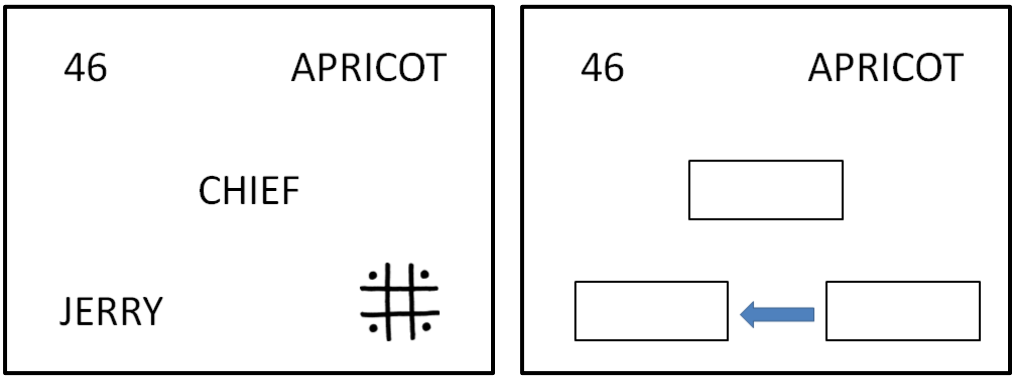


Figure 3.8: Example of arrays used. The image on the left shows the array as presented at encoding. The image on the right shows the array as presented at recall and FOK, and at recognition. The target item is indicated by the blue arrow.

3.4.3 Results

Due to the complexity of the symbols, it was decided that half marks could be awarded for attempts which correctly represented the key features of the symbol. For example, a half mark would be given if the participant recorded the hash design from the example in Figure 3.8 above but failed to reproduce the dots. Responses were scored as fully correct if all features of the symbol were present. For numbers, fruit/vegetables, names and five letter words only full responses were considered correct.

3.4.3.1 Memory

Recall and recognition responses were analysed first with a 2 (age: young adults and older adults) x 2 (cues: 2 cues or 4 cues) repeated measures ANOVA (see Table 3.8). Proportions were calculated by dividing the number of correct responses by the total possible responses for the condition (i.e. 15). The proportion of correct recall responses given showed no effect of age, $F(1,59) = 2.752$, $p = .102$, $\eta_p^2 = .045$. A main effect of cues was present, $F(1,59) = 7.919$, $p = .007$, $\eta_p^2 = .118$, with greater recall accuracy when 4 cues were given than when 2 cues were given, as expected. No interaction was found between age and cue, $F<1$, thus young and older adults benefited equally from the provision of cues. Recognition performance showed a similar pattern, with no effect of age, $F<1$, but the expected main effect of cues, $F(1,59) = 5.484$, $p = .023$, $\eta_p^2 = .085$. Thus the extra support given by four cues being present at retrieval leads to higher memory performance than when only two cues are given. Again, no interaction between age and cues was present, $F<1$.

Table 3.8: Proportion of items correctly recalled and proportion of items correctly recognised by young and older adults

	Young adults		**Older adults**	
	2 cues	**4 cues**	**2 cues**	**4 cues**
Recall	0.070 (.077)	0.096 (.108)	0.032 (.061)	0.062 (.085)
Recognition	0.668 (.139)	0.698 (.156)	0.640 (.189)	0.703 (.136)

Commission errors were also examined. Older adults gave no commission errors when presented with either 2 cues or 4 cues; thus older adults either correctly recalled information or did not report any information, indicating a conservative threshold for retrieval. Consequently, only young adults were subject to further analysis to examine whether an effect of cues was present. A paired samples t-test revealed that the number of cues available at recall had no impact on the proportion of commission errors observed, $t(40) = 1.305$, $p = .199$.

The 2 cue condition also provided the opportunity to assess partially retrieved information. Participants were asked to report any missing details from the array other than the identified target item. The number of correct and incorrect responses given were examined using a 2 (age) x 2 (accuracy: correct or incorrect) mixed ANOVA (see Table 3.9). No effect of age was present, $F(1,59) = 2.734$, $p = .104$, $\eta_p^2 = .044$, indicating that young and older adults retrieved a similar amount of partial information. Surprisingly, no effect of accuracy was found, $F<1$; participants were equally likely to report correct information as incorrect information. This could suggest a fairly liberal response criterion. Finally, no interaction was observed between age and accuracy, $F(1,59) = 1.760$, $p = .190$, $\eta_p^2 = .029$. It should be noted that a total of 24 young adults and 9 older adults reported any partial information. Due to the low reporting rate this data was not subjected to further analysis.

Table 3.9: The accuracy of partial information reported during the 2 cue condition for young and older adults

	Young adults	**Older adults**
	M (SD)	**M (SD)**
Partial information: Correct	1.305 (2.024)	0.350 (0.489)
Partial information: Incorrect	0.671 (1.709)	0.500 (1.395)

3.4.3.2 Metacognition

The magnitude of ratings for unrecalled items was examined first, again with a 2 (age) x 2 (cues) ANOVA (Figure 3.9). A main effect of age was found $F(1,59) = 18.603$, $p < .000$, $\eta_p^2 = .240$, with young adults giving higher FOK ratings than older adults. A main effect of cues was also present, $F(1,59) = 26.940$, $p < .001$, $\eta_p^2 =$

.313, with FOK ratings being higher when 4 cues were given at recall than when 2 cues were given. The interaction between age and cue was nonsignificant, *F*<1.

The ratings for items subsequently recognised were compared to those which were not recognised using a 2 (age) x 2 (cue) x 2 (status: recognised or not recognised) ANOVA (see Figure 3.9). A main effect of age was again present, $F(1,59) = 17.399$, $p < .001$, $\eta_p^2 = .228$, with young adults giving higher ratings than older adults, and a main effect of cues, $F(1,59) = 19.771$, $p < .001$, $\eta_p^2 = .251$, with 4 cues leading to higher ratings than 2 cues. Importantly, a main effect of status was also found, $F(1,59) = 5.613$, $p = .021$, $\eta_p^2 = .087$. Unrecalled items which were subsequently recognised received higher FOK ratings than unrecalled items which were not subsequently recognised. Participants therefore show some level of accuracy in their FOK predictions, as their ratings do discriminate between future memory performance. No interactions reached significance (age and cues: *F*<1; age and status: *F*<1; cues and status: $F(1,59) = 1.936$, $p = .169$, $\eta_p^2 = .032$; age, cues and status: *F*<1).

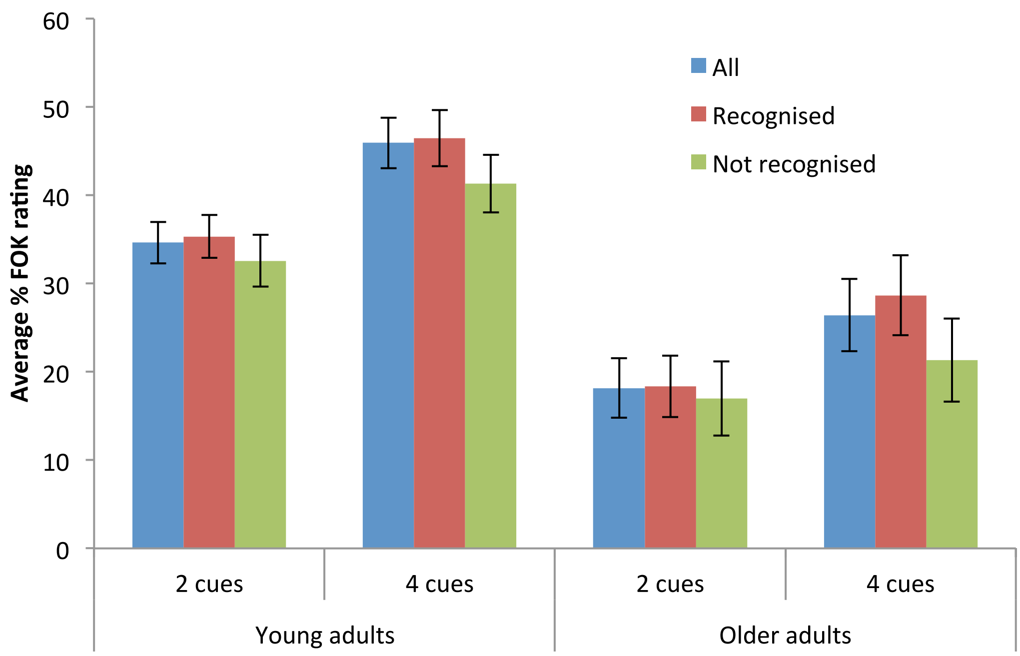


Figure 3.9: Magnitude of FOK ratings given by young and older adults on scale of 0 to 100%. Ratings are presented for all unrecalled items regardless of subsequent recognition, and also according to whether unrecalled items were subsequently recognised or not.

A gamma correlation was calculated between the different rating categories and subsequent recognition (see Table 3.10). Although the previous analysis indicated a level of predictive accuracy to participants FOK judgements, none of the gamma correlations were significantly above chance (young adults 2 cues: $t(39) = 1.751$, $p = .088$; young adults 4 cues: $t(39) = 0.172$, $p = .864$; older adults 2 cues $t(18) = 0.074$, $p = .942$; older adults 4 cues: $t(19) = 1.318$, $p = .203$). When considered using a 2 (age) x 2 (cues) ANOVA, main effects of age, $F<1$, and of cues, $F<1$, failed to reach significance. However, a marginal interaction was present, $F(1,56) = 3.009$, $p = .088$, $\eta_p^2 = .051$. Although follow up t-tests failed to reach significance, the interaction would appear to be driven by older adults providing higher accuracy FOK judgements when given 4 cues, whereas young adults provide less accurate FOK predictions when 4 cues are given compared to 2 cues.

Table 3.10: Gamma correlations for young and older adults when given two cues at retrieval and when given four cues at retrieval

	Young adults		**Older adults**	
	2 cues	**4 cues**	**2 cues**	**4 cues**
Gamma	0.133 (.49)	-0.008 (.62)	0.011 (.62)	0.227 (.64)

3.4.4 Discussion

The present study aimed to examine the effect of environmental support at retrieval on FOK accuracy in ageing. By asking participants to learn multiple items together it was possible to vary the number of cues given when eliciting a target response, thus varying the amount of environmental support available at recall, FOK and recognition. Memory accuracy showed a clear effect of support, with a greater number of cues leading to greater accuracy in both recall and recognition measures. It should be noted that recall performance was particularly low on this task, suggesting both young and older adults had difficulty recalling target items. Indeed, no effect of age was observed on either recall or recognition. However, recognition accuracy was of an acceptable level, suggesting participants had sufficiently encoded the items and should therefore be able to access diagnostic partial information on which to make an FOK judgement. The equivalent memory performance in young and older adults allows further examination of the potential effect of age on FOK without the confound of differing memory accuracy.

With regards to the FOK, judgements of future recognition are quite low, with average ratings below 50%. This again demonstrates the difficulty of the task generally. A clear age effect was also shown, with older adults being particularly underconfident with their ability to later recognise unrecalled items compared to young adults. Despite similar levels of actual memory performance, the older adults perceive the task as being especially difficult and reduce their ratings accordingly. The pattern of ratings does mirror that of memory accuracy, with both young and older adults giving higher ratings for items when four cues are presented at recall than when two cues are presented. Thus environmental support does affect the magnitude of the FOK rating given, with more cues leading to a higher confidence in future recognition.

With regards to FOK accuracy, there are some indications that participants are able to discriminate between items which will be subsequently recognised and those which will not. Unrecalled items which are later recognised reliably received ratings which were higher than those items which were not recognised. However, no interaction with cues was found, suggesting that providing more support at retrieval did not affect how well young and older adults were able to differentiate future recognition ability. Environmental support would appear to affect the magnitude of the ratings without affecting their accuracy. Gamma correlations were also calculated to ascertain FOK accuracy, however correlations failed to reach above chance accuracy in any condition. Consequently, no effects of age or of cues were detected. The gamma correlation gives a more robust measure of FOK accuracy than comparing ratings for recognised and unrecognised items (Nelson, 1984). Further consideration of the effect sizes for the gamma analysis combined with the low level of recall accuracy does however suggest that the present task may be too difficult for participants to be able to monitor their memory sufficiently to report their subjective feelings. Therefore it is premature to conclude that environmental support at retrieval does not have an influence on FOK accuracy. A revised task where recall accuracy was of a higher level may be able to detect the effects of cue manipulation on FOK accuracy.

A second experiment examining the effect of environmental support was conducted concurrently to Experiment 3.3. This allowed a second comparison of the effect of two cues or four cues provided at retrieval. In addition, the recognition task was altered to consider the impact of the associative deficit in ageing.

3.5 Experiment 3.4: Ageing and environmental support at retrieval II

3.5.1 Introduction

The Associative Deficit Hypothesis (ADH; Chalfonte & Johnson, 1996; Naveh-Benjamin, 2000) proposes that the effects of ageing on episodic memory performance are due to deficits in encoding and retrieving associations between stimuli. Indeed, recognition of individual items in older adults has been shown to be greater than recognition of associations with various types of stimuli (see Old & Naveh-Benjamin, 2008 for a meta-analysis). The FOK task necessarily involves associative encoding, as a cue is required to elicit a specific memory and gain a specific FOK when recall is not accomplished. The associative nature of the recognition test can, however, be manipulated by using a forced choice task with foils being either novel items or recombined pairs from the encoding stage. Both types of recognition task have been used extensively in the metacognitive literature.

Experiment 3.3 considered the effect of environmental support at retrieval on FOK magnitude and accuracy, finding that as more support was provided FOK ratings increased in magnitude and there were also some indications of improvement in FOK accuracy. Experiment 3.4 was conducted concurrently with Experiment 3.3, with an adjustment to the recognition task[4]. Experiment 3.4 used a parallel form of the previous experiment, with a different set of participants asked to learn the arrays. The number of cues available at recall, FOK and recognition was also manipulated in the same fashion as the previous study, with the target accompanied by either two cues only or four cues. A four alternative forced choice task was again used, however the foils were items which had been encoded in association with other targets, thus creating an associative recognition task. If ADH is correct, older adults should show a deficit in recognition accuracy compared to young adults. The supportive effects of increased cues at retrieval can also be examined in relation to an associative recognition task, as increased environmental support may mitigate the effects of the associative recognition task in the older adult group.

3.5.2 Method

3.5.2.1 Participants

Participants were 41 A-level and undergraduate Psychology students (age range 16-27, *M* = 18.34, *SD* = 2.40) and 20 older adults (age range 64-85, *M* = 73.9, *SD* = 6.87). Recruitment, screening and ethical procedures were identical to those in Experiment 3.3 (older adults MMSE scores *M* = 28.9, *SD* = 1.21).

3.5.2.2 Materials

Materials were identical to those used in Experiment 3.3, except for in the Recognition test. For this study, distracter items were taken from the pool of target items only. Items from each array occurred an equal number of times as distracters so that priming of certain arrays did not occur. The answer booklets were constructed so that distracters were not identified on adjacent arrays, therefore

[4] Due to differences in recall accuracy between young adults over Experiments 3.3 and 3.4 the data could not be collapsed across experiments and so is presented independently

participants could not use the information presented in the booklet to eliminate distracters.

3.5.2.3 Procedure

The procedure for this study was identical to that of Experiment 3.3. Participants were tested in small groups, were allowed 12 seconds to encode each slide, and arrays were presented in two blocks with recall, FOK and recognition occurring for each stage.

3.5.3 Results

3.5.3.1 Memory

Analysis and calculation of proportions follows that of Experiment 3.3. Recall and recognition were analysed as in the previous experiment, with 2 (age) x 2 (cues) ANOVAs. A main effect of age was found on recall accuracy, $F(1,59) = 11.179$, $p = .001$, $\eta_p^2 = .159$, with young adults recalling more items than older adults (see Table 3.11). A main effect of cues was also present, $F(1,59) = 5.261$, $p = .025$, $\eta_p^2 = .082$, with recall higher when 4 cues were provided than when 2 cues were given, as in the previous study. No interaction between age and cues was present, $F<1$. Recognition accuracy followed the same pattern, with young adults recognising significantly more items than older adults, $F(1,59) = 19.605$, $p < .001$, $\eta_p^2 = .249$, and 4 cues leading to higher accuracy than 2 cues, $F(1,59) = 15.002$, $p < .001$, $\eta_p^2 = .203$. Again, no interaction between age and cues was present for recognition accuracy, $F<1$.

Table 3.11: Proportion of items correctly recalled and proportion of items correctly recognised by young and older adults

	Young adults		**Older adults**	
	2 cues	**4 cues**	**2 cues**	**4 cues**
Recall	0.121 (.107)	0.166 (.156)	0.032 (.049)	0.067 (.118)
Recognition	0.468 (0.165)	0.574 (.176)	0.303 (.163)	0.390 (.179)

Both young (2 cues: M = 0.072, SD = 0.090; 4 cues: M = 0.086, SD = 0.098) and older adults (2 cues: M = 0.053, SD = 0.093; 4 cues: M = 0.083, SD = 0.133) made errors of commission in the current study. These were examined with a 2 (age) x 2 (cues) mixed ANOVA. No main effect of age was found, F<1. The effect of cue was marginally significant, $F(1,59) = 2.911$, $p = .093$, $\eta_p^2 = .047$, with slightly more errors reported for the 4 cue condition than the 2 cue condition. No interaction was present, F<1.

As with Experiment 3.3, partial information was also examined with a 2 (age) x 2 (accuracy) mixed ANOVA (see Table 3.12). A main effect of age was observed, $F(1,59) = 4.363$, $p = .041$, $\eta_p^2 = .069$, with young adults providing more partial information than older adults. The main effect of accuracy was marginally significant, $F(1,59) = 3.852$, $p = .054$, $\eta_p^2 = .061$, with accurate responses given slightly more than inaccurate responses. An interaction between age and accuracy was also present, $F(1,59) = 6.061$, $p = .017$, eta = .093. Follow up t-tests revealed that young adults gave more correct responses than older adults, $t(59) = 2.877$, $p = .006$, while both age groups gave a similar number of incorrect responses, $t(59) = 0.012$, $p = .991$. Furthermore, while young adults gave significantly more correct responses than incorrect responses, $t(40) = 3.828$, $p < .001$, older adults gave a similar number of correct and incorrect responses, $t(19) = 0.311$, $p = .759$. Thus older adults were less discriminating in their criterion for responding, leading to information being equally likely to be correct as incorrect. A total of 31 young adults and 9 older adults provided partial information. Due to the low number of older adults who were able to report these details, partial information data were not subjected to further analysis.

Table 3.12: The accuracy of partial information reported during the 2 cue condition by young and older adults

	Young adults	**Older adults**
	M (SD)	***M (SD)***
Partial information: Correct	2.134 (2.101)	0.650 (1.348)
Partial information: Incorrect	0.805 (1.364)	0.800 (1.765)

3.5.3.2 Metacognition

As previously, the magnitude of ratings for all unrecalled items was analysed with a 2 (age) x 2 (cues) ANOVA (see Figure 3.10). A main effect of age was found, $F(1,59) = 21.854$, $p < .001$, $\eta_p^2 = .270$, with young adults giving higher ratings than older adults. A main effect of cues was also present, $F(1,59) = 19.879$, $p < .001$, $\eta_p^2 = .252$. Ratings were higher for unrecalled items prompted with 4 cues than those with 2 cues. No interaction was found, $F<1$. To ascertain accuracy, ratings for unrecalled items which were subsequently recognised were compared to unrecalled items which were not subsequently recognised with a 2 (age) x 2 (cues) x 2 (status) ANOVA (see Figure 3.10). Ratings given by older adults were significantly lower than those of young adults, $F(1,59) = 19.183$, $p < .001$, $\eta_p^2 = .245$, and 4 cues elicited higher FOK ratings than 2 cues, $F(1,59) = 25.326$, $p < .001$, $\eta_p^2 = .300$. The crucial main effect of status was also significant, $F(1,59) = 10.005$, $p = .002$, $\eta_p^2 = .145$, with items later recognised receiving higher ratings than those that were not later recognised. No interactions were present (age and cues: $F<1$; age and status: $F<1$; cues and status: $F(1,59) = 2.531$, $p = .117$, $\eta_p^2 = .041$; age, cues and status: $F<1$).

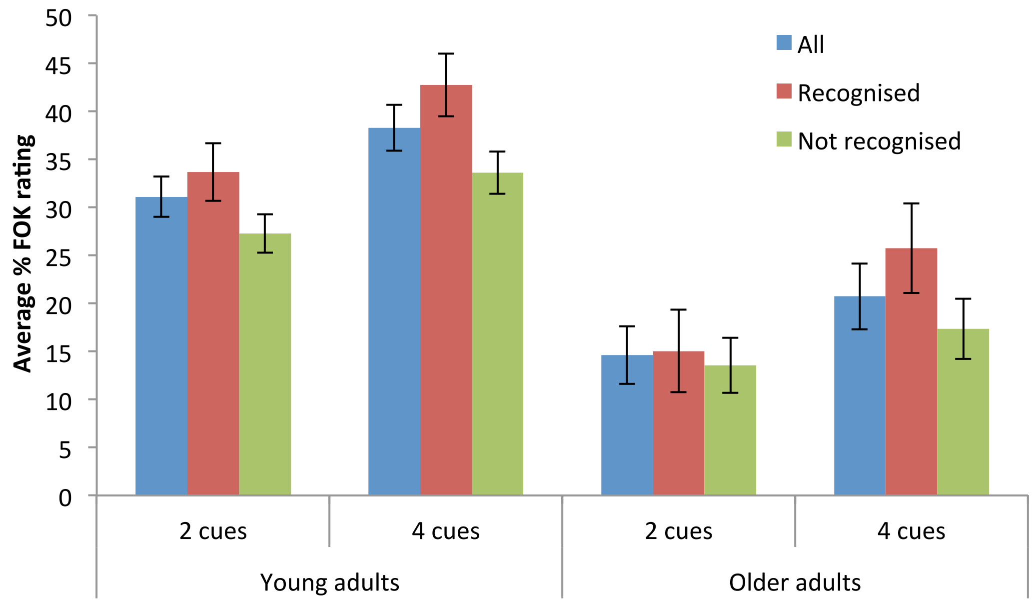


Figure 3.10: Magnitude of FOK ratings given by young and older adults on scale of 0 to 100%. Ratings are presented for all items regardless of subsequent recognition, and also according to whether items were subsequently recognised or not.

Gamma correlations (see Table 3.13) indicated that FOK accuracy was at chance levels in the young adult group for both the 2 cue, $t(40) = 0.344$, $p = .733$, and the 4 cue conditions, $t(40) = 1.351$, $p = .184$. Older adults were also at chance for the 2 cue condition, $t(17) = 0.934$, $p = .364$. However for the 4 cue condition FOK accuracy was found to be above chance, $t(17) = 2.325$, $p = .033$, indicating that older adults were able to accurately predict future recognition of items when higher support was available at retrieval. A 2 (age) x 2 (cues) ANOVA revealed no effect of age on gamma correlations, $F<1$. A main effect of cues was present, $F(1,57) = 5.560$, $p = .022$, $\eta_p^2 = .089$, with 4 cues leading to higher accuracy of FOK ratings than 2 cues. No interaction between age and cues was present, $F(1,57) = 2.648$, $p = .109$, $\eta_p^2 = .044$.

Table 3.13: Gamma correlations for young and older adults when given two cues at retrieval and when given four cues at retrieval

	Young adults		**Older adults**	
	2 cues	**4 cues**	**2 cues**	**4 cues**
Gamma	0.029 (.54)	0.109 (.52)	-0.159 (.72)	0.278 (.51)

3.5.4 Discussion

The current study replicated Experiment 3.3. In contrast to Experiment 3.3, the present study found a clear effect of age on recall performance as well as recognition performance. The age effect on recognition accuracy is to be expected from the predictions of ADH, in that older adults find associative recognition tasks more difficult and should therefore exhibit lower accuracy than young adults (Naveh-Benjamin, 2000). This explanation is somewhat called into question for the present study however due to the lower recall accuracy in older adults. The low recall accuracy suggests a more general memory deficit in the older adult group, thus the age effect on recognition could be due to a global deficit in encoding affecting all measures of memory rather than the presence of associative foils. Why an effect of age is shown in the present study when none is found in Experiment

3.3 is unclear, as the procedure and materials used up to the point of eliciting recall and FOK judgements are identical in the two studies. Nonetheless, an age effect on recall is clearly found in the present study.

The pattern of results is in agreement between the two studies. The manipulation of environmental support at retrieval again led to higher FOK ratings when more support was given than when less support was given. Improved support not only allows for greater explicit retrieval but also allows access to more partial information for unrecalled items, resulting in higher ratings. Older adults remained less confident than young adults in their future recognition, and both age groups gave higher ratings to items which were subsequently recognised than items which were not subsequently recognised. The provision of more cues at retrieval did not allow participants to improve their ability to discriminate between future recognition states.

Interestingly, despite deficits in memory ability, gamma correlations for the older adult group when given four cues at retrieval did reach above chance levels, further indicating a level of diagnostic accuracy in their ratings. Young adults however failed to reach above chance accuracy in any of their FOK predictions. Although there was no main effect of age, this slight advantage in FOK accuracy for the older adults is unusual. As described in Chapter 2, age effects on the FOK are proposed to be in the direction of a disadvantage in older participants, whether these effects are due to a reduced ability to encode information (MacLaverty & Hertzog, 2009; Perfect & Stollery, 1993) or due to an impairment in memory monitoring itself (Souchay et al., 2007; Thomas et al., 2011). Both of these explanations of FOK deficits in ageing suggest that deficits in FOK accuracy should have been found in the older adult group within the current study, as an age effect was found on memory performance thus reducing encoding ability and the diagnostic accuracy of the partial information older adults were able to retrieve. The absence of this deficit could be due to the environmental support available. As already noted, the environmental support hypothesis (Craik 1983, 1986) proposes that, when given sufficient support, memory accuracy in older adults can improve up to a similar level as that of young adults. Although this effect on memory did not occur in the present study, the cues available may have enabled older adults to discern future recognition ability to a greater extent than in typical FOK studies where a single cue is given. This explanation is tentative and it is beyond the scope of this chapter to further examine, but it does provide a potential avenue for future research.

The effect of environmental support at retrieval clearly influenced memory and FOK magnitude measures. In the present study, it also influenced the accuracy of the FOK ratings both when considering the magnitude of ratings by future recognition status and when considering the gamma correlation. Koriat (1993) observed that as the amount of partial information participants could retrieve increased, so did the magnitude of their FOK ratings. In addition, Thomas et al. (2011) found that by prompting older adult participants in particular to retrieve partial information prior to making an FOK judgement, the accuracy of the judgements increased. Thus the partial information a participant is able to retrieve impacts both the magnitude and accuracy of the FOK. The present study has shown that, in addition to the partial information retrieved, the partial information provided when making an FOK also influences the magnitude and accuracy of the FOK. Importantly, this effect occurred independent of age effects. Both young and older adults gave higher FOK ratings when more cues were provided, and achieved greater gamma accuracy when more cues were provided.

3.6 General discussion

The present chapter aimed to examine the FOK with manipulations of retention and retrieval. Studies have typically focused on effects induced at encoding, using paradigms with known impacts on memory accuracy and examining their influence on metacognitive accuracy. However, there are also numerous classic effects on memory that occur through manipulations at retention and retrieval which may be useful for examining FOK. Experiments 3.1 and 3.2 examined how the placement of a retention interval within the FOK paradigm might affect reliance on different types of metacognitive information, and its effect within ageing. Experiments 3.3

and 3.4 examined the effect of environmental support at retrieval and potential differences in young and older adults.

The manipulation of retention interval examined in Experiment 3.1 demonstrated that, similar to JOLs, FOKs can be influenced by both theory and experience based judgements (Koriat et al., 2004). By placing the delay either before or after the FOK was obtained, it was possible to change the emphasis that participants placed on each type of judgement, thereby affecting the magnitude of the FOK. However, no impact of delay was observed for FOK accuracy. Irrespective of the placement of the delay, or indeed the presence of the delay, gamma correlations were found to be equivalent between the three experimental conditions. This differential effect of delay on FOK magnitude and accuracy was replicated with Experiment 3.2, whereby both young and older adults showed reduced FOK ratings after the delay with no effect on accuracy.

As described in Chapter 2, the accessibility account and the Memory Constraint Hypothesis (MCH) have both been proposed to explain how FOKs are formed. Although the MCH was primarily developed to address observed differences in FOK accuracy in ageing, it can logically be extended to provide a more general account of FOK functioning. The accessibility account (Koriat, 1993, 1997) suggests that FOKs are based on the quantity and quality of the partial information available, with more partial information leading to higher and more accurate FOKs. The MCH (MacLaverty & Hertzog, 2009) suggests that the accuracy of FOKs is influenced primarily by the level of encoding, with better encoding leading to more accurate FOKs. The findings of Experiments 3.1 and 3.2 thus have interesting implications for each theory of FOK. The MCH predicts no effect of delay on FOK, as all participants experienced the same encoding conditions, thus should exhibit equivalent FOK accuracy. The accessibility account predicts an effect of delay, as forgetting will have occurred thus reducing the quality and quantity of partial information available when making the FOK. Findings of Experiment 3.1 are primarily consistent with the MCH account, with no main effects of delay shown. However, findings from Experiment 3.2 are not. Older adults showed lower memory accuracy than young adults, and yet their FOK accuracy was similar to that of the young adults. In addition, nuances within the data of both experiments, with gamma accuracies being above or at chance level, cannot be explained by the MCH alone.

The accessibility account can explain the differences the MCH fails to. As previously described, the partial information available upon a failed recall attempt may be experience based or theory based. The placement of delay in Experiments 3.1 and 3.2 varies the reliance on each type of judgements. Those which are theory based are more prone to error, thus reducing their accuracy. The lack of an age effect in Experiment 3.2 can also be explained when combined with the observation of reduced self-initiated strategy use in older adults. By not spontaneously adjusting for the effect of delay, older adults may have consequently increased their FOK accuracy compared to young adults.

The manipulation of retrieval conditions also offers some new insights into the two competing theories of FOK. Experiments 3.3 and 3.4 demonstrated the impact of providing partial information at retrieval, thus reducing the demands on self-initiated processing to make accurate FOK judgements. Experiment 3.3 observed no effect of age on memory performance and no effect of age on FOK accuracy, as the MCH would predict. In addition, the level of environmental support did impact FOK accuracy, with more cues provided at recall and FOK leading to greater FOK accuracy, as both the MCH and the accessibility account would predict. Experiment 3.4 is somewhat more problematic for the MCH. In this instance, an effect of age on memory performance was observed, with young adults recalling more items than older adults. And yet, no age effect was found on gamma correlations. The accessibility account does provide a possible solution to this finding. As the partial information is provided, older adults are not required to spontaneously retrieve it, thus it can be used to inform their judgements. This extra environmental support may be all that is needed for older adults to refine their predictive abilities, as shown previously by Thomas et al. (2011), thus removing the potential deficit to FOK caused by reducing encoding ability.

The MCH was originally conceptualised to explain the effect of age on FOK accuracy. Its proposition that reduced FOK accuracy is fundamentally due to reduced encoding efficiency implies a close link between memory function and FOK accuracy. In many instances this more direct association between memory and FOK is evident, with manipulations of encoding also affecting predictive accuracy. However, encoding is only one aspect of a memory. Retention and retrieval are also key components that can affect the retrievability of information. As demonstrated in the present chapter, changes to the task demands made during retention and retrieval can also influence FOK accuracy. Thus the MCH, while being an extremely useful explanation of the potential task effects on FOK, is incomplete. Likewise, the accessibility account is not, in isolation, sufficient to explain all aspects of data presented here. It may therefore be more appropriate to see these theories as two complementary frameworks on which the FOK is made, and the contribution that each mechanism makes towards the FOK judgement may be determined by the constraints of the task. An important step to further understand the potential interplay between these two theories is to go beyond manipulations of encoding to further explore other memory phenomena during consolidation and retrieval.

The observed floor effects in recall for Experiments 3.3 and 3.4 could be considered as limiting the interpretation of these data. Low levels of recall could suggest the arrays were not fully encoded and this may have limited the retrieval of partial information on which the FOK could be based, reducing participants’ ability to make metacognitive predictions. However a number of observations suggest that, despite the low recall levels in both young and older adults, the results are nonetheless a valuable contribution to the FOK theory. Firstly, recognition accuracy remains high, with chance performance being 0.25 and observed accuracy being significantly above this value for all age groups across both studies (see Appendix B). Thus the arrays have been sufficiently encoded to form some representation in memory, even though it is not available for explicit recall. Likewise, the expected age effect was observed on recall in Experiment 3.4, again suggesting encoding was successful to some extent. Finally, the sensitivity of FOK judgements to the amount of environmental support at retrieval is difficult to account for if task difficulty had impaired recall accuracy to such an extent that metacognitive judgements could not be made. In sum, although floor effects were observed within these experiments, the finding suggest that encoding was sufficient to allow for memory formation, enabling acceptable levels of recognition performance and providing diagnostic information on which the FOK could be made. It would therefore be recommended that the task be replicated to further support the observations in sensitivity of FOK magnitude and the suggestion of an effect within FOK accuracy.

Chapter 4. Development of a Signal Detection Model of Feeling of Knowing

4.1 General introduction

The FOK paradigm has yielded over 100 scientific articles[5] and given rise to several models of metacognitive function. Interestingly, whilst many mathematical and computational models of recognition memory exist, there has been limited application of these types of model within the FOK. Consideration of these types of models has the potential to not only further our theoretical understanding of the FOK but could also provide novel avenues of research and a more robust measurement of FOK accuracy. Following an influential paper by Nelson (1984), the Goodman-Kruskal gamma coefficient has become the predominant measure of accuracy within the metacognitive literature. More recently, it has been proposed that signal detection theory could provide a more reliable form of measurement (Masson & Rotello, 2009). The current chapter explores some of the limitations of the Goodman-Kruskal gamma correlation and their implications for FOK research. It will then consider the signal detection models of recognition memory and the issues of adopting these for the FOK, as well as considering the limitations of the model proposed by Masson & Rotello (2009). It then proposes a novel model of FOK based on signal detection theory, with tests of the model using both simulated data and experimental data from Chapter 2. Discussion of the model outcomes and potential future exploration of FOK within the signal detection theory framework are then considered.

[5] Based on a search of Web of Science on 6/9/2013 using the keywords 'FOK' and 'Memory'.

4.1.1 Accuracy and the gamma statistic

Accuracy can be examined as either absolute accuracy or relative accuracy. Absolute accuracy, or calibration, is the degree to which judgements and observed events agree. For example, if a participant predicts they will recognise 60% of unrecalled items, perfect calibration would be shown if at test they recognised 60% of the items. Recognising a smaller or larger proportion would reflect inaccurate judgements, and the level of over- or underconfidence can be established. However, this conceptualisation of accuracy misses one of the key features of the FOK, the idea that there is a relationship between the two variables being measured. FOK accuracy relies on both variables, thus it is the relationship between feeling of knowing predictions and levels of recognition on an item by item basis that is of primary interest. This is termed relative accuracy, or resolution, and is the concept of accuracy used throughout this thesis and indeed throughout the majority of the metacognitive literature.

Accuracy of the FOK has been predominantly measured using the Goodman-Kruskal gamma correlation (Goodman & Kruskal, 1954; Nelson, 1984), subsequently referred to simply as gamma. Nelson identified six characteristics which were desirable in a measure of FOK accuracy. These include that FOK accuracy should be independent of the type of criterion test, the difficulty of the criterion test, the level of performance on the criterion test, and that as FOK ability increases the accuracy score should also increase. Of the eight measures that had previously been used in the literature, Nelson concluded that the gamma correlation upheld the six desirable properties required and was therefore the most appropriate accuracy measure available for FOK researchers.

Table 4.1 shows the possible observations of the standard Hart feeling of knowing methodology. Here, a simple Yes/No FOK is made, with subsequent recognition either being successful or unsuccessful. This leads to four independent outcomes which may occur for any item. Outcome *a* results from a positive FOK prediction accompanied by successful recognition; the person predicts that they will recognise the item later and they do indeed recognise it. Outcome *b* occurs if a person predicts they will recognise the item later but fails to identify it at recognition. Outcome *c* corresponds to a negative FOK but successful recognition, and outcome *d* is a result of a negative FOK prediction and no recognition occurring. Outcomes *a* and *d* thus reflect accurate FOK predictions whereas outcomes *b* and *c* reflect inaccurate FOK predictions. To use similar terminology to that used in recognition memory, *a* is a hit, *b* is a false alarm, *c* is a miss, and *d* a correct rejection.

Table 4.1: General 2 x 2 array for Feeling of Knowing judgements following Hart's methodology

	Recognition		
Feeling of Knowing	**Yes**	**No**	**Σ**
Yes	*a*	*b*	*a + b*
No	*c*	*d*	*c + d*
Σ	*a + c*	*b + d*	*n = a + b + c + d*

Gamma is calculated by comparing the outcome for each item with the outcome obtained on each other item, then adding up the number of concordant and discordant pairings. Concordant pairings occur when item A is judged to have a higher FOK than item B, and item A is subsequently recognised whereas item B is not. The number of concordant pairings can be obtained by multiplying outcome *a* by outcome *d* in the present array. A discordant pairing occurs when item A receives a higher FOK rating than item B, but item B is recognised whereas item A is not. Again, the number of discordant pairings can be calculated by multiplying outcome *b* by outcome *c* in Table 4.1. Gamma is therefore a measure of relative accuracy, and is calculated as follows:

$$\boldsymbol{\gamma = \frac{ad - bc}{ad + bc}} \qquad \textbf{(1)}$$

Note that pairs of items given identical FOK ratings or with identical outcomes are not be included in the calculation of gamma.

Despite general acceptance and extensive use of the gamma correlation within the FOK literature, not all researchers agree that gamma is the most appropriate measure of FOK accuracy. In particular, Schraw (1995) argued that a more complete examination of FOK accuracy would be to consider not only gamma but also the Hamann coefficient (HC). While gamma is a multiplicative model, HC is additive:

$$\boldsymbol{HC = \frac{(a+d) - (b+c)}{(a+d) + (b+c)}}$$

(2)

This mathematical difference has associated implications. The additive nature of HC means that it is margin dependent, which is not a suitable property for a measure of FOK to have. By being dependent on the marginal values, this constrains the total value which the measure can take. The greater the difference between the marginals for FOK and the marginals for recognition, the lower the upper limit HC, or indeed any other margin dependent measure, can achieve (Bishop, Fienberg, & Holland, 1975; Nelson, 1984). Thus the maximum range of the measure will vary for each participant. Gamma, meanwhile, is capable of reaching all values from -1 to +1, irrespective of the marginal values. A further constraint of HC is that it cannot be used in situations where the two variables are not equal in size. Although the focus of this chapter is the simple 2 x 2 paradigm originally created by Hart (1965), more complex FOK ratings can be obtained than the original Yes or No decision. Indeed, although Chapter 2 focuses in the Yes/No FOK paradigm, Chapter 3 used a rating scale with participants able to rate their FOK from 0% to 100%. These multiple response options for participants when making FOK judgements are in fact the most common in the FOK literature. The gamma correlation copes with this easily, therefore the same measure in the 2 x 2 case can be applied to a n x 2 case or, if necessary, a n x m case. HC however is limited to the 2 x 2 instance only. Thus it can only be of use in a very restricted set of circumstances; a second measure would be needed to encompass all previous findings whereby uneven dimensions occur.

4.1.2 Limitations and considerations when using gamma

Gamma has been demonstrated to be of great utility in measuring relative accuracy compared to other measures, and yet there are still a number of potential issues when using gamma to examine accuracy. One such issue which has been raised is a lack of stability across alternate test forms and test halves (Nelson, 1988; Thompson & Mason, 1996). Although the FOK has been shown to have high test-retest reliability when the same items are presented twice (Costermans, Lories, & Ansay, 1992; Nelson et al., 1984, 1986), when different items are used at each test this reliability disappears. For example, Nelson (1988) obtained FOK ratings for 110 unrecalled general knowledge questions. A separate gamma score was calculated for odd items and even items for each participant, allowing a Spearman's rho correlation to be calculated to assess split-half stability. However, this stability coefficient only reached -.02. Similarly low levels of stability were shown when fewer test items were used. Thompson & Mason (1996) further explored the stability of gamma with confidence judgements. In a series of three experiments, they demonstrated that gamma showed low split-half and alternate forms stability for three different memory tasks for young adults, older adults, and when recall or recognition was used as the criterion test. For the 13 assessments of split-half reliability the average stability coefficient was .15, while the average for the six assessments of alternate forms stability coefficient was also very low at .02. This lack of stability is a point of concern for studies interested in the individual differences in metacognitive accuracy. If the measure is not stable across tests for each individual, it is not possible to accurately assess the impact of experimental manipulations on accuracy on a person by person basis.

It should be noted that, equally, the underlying construct may vary. This would imply that for each individual item participants would set a new level at which they decided whether items will or will not be recognised. Although this is plausible, and indeed likely that some variation in criterion placement occurs (see Benjamin, Diaz, & Wee, 2009; Mueller & Weidemann, 2008 for a similar argument within recognition memory models), the variation between each item would need to be large to reduce the split-half correlations of gamma to chance levels. This level of variance would be cognitively expensive. Therefore, in the absence of any experimental evidence to the contrary, it is assumed that participants have a consistent criterion across items, which may show a small variation but not to the extent that accuracy will be affected across subsets of items.

Although this stability issue does present a problem for certain investigations of FOK and metacognition more generally, the vast majority of studies are more concerned with group effects of single experimental manipulations. For example, the experiments presented in the present thesis have been concerned with the effects of delay (Experiments 3.1 and 3.2), availability of cues at test (Experiments 3.3 and 3.4), episodic and semantic effects (Experiments 2.1 and 2.2) and ageing (Experiments 2.1, 2.2, 3.2, 3.3 and 3.4). Those experiments where individuals were required to make multiple FOK judgements have used the same items at each test. In these instances, where manipulation effects are examined on a group level and the same items are used at each test, gamma remains reliable and is, as proposed by Nelson (1984), the most appropriate measure of FOK accuracy and no caveats need be placed on interpretation of the accuracy measure. Researchers who wish to explore FOK on a more individual basis would however need to consider how to design a novel methodology to encourage metacognitive stability and therefore allow a focus on individual differences (Nelson, 1988; Thompson & Mason, 1996).

A second consideration when using gamma as a measure of FOK accuracy is the number of alternates given in the criterion test (Leonesio & Nelson, 1990; Schwartz & Metcalfe, 1994). Consider when the participant does not know the correct answer, and so gives a low FOK rating. At test, it would be expected that they will choose an incorrect answer, demonstrating accurate metacognition. However, the likelihood of choosing an incorrect answer varies with the number of alternative to choose from. With a Yes-No recognition task, the participant has a 50% chance of choosing the correct response despite having no memory of what the correct answer is. In contrast, when four alternatives are presented, the chance of choosing the correct answer is only 25%. As the number of alternatives increases, the chance of choosing the correct answer decreases. This in turn reduces the likelihood that, when a participant has no memory for the correct target and gives a low FOK judgement, they will by chance choose the appropriate response, thereby appearing as though they have given an inaccurate FOK judgement. Subsequently, as more alternatives are available at criterion test, the level of accuracy measured by gamma will also increase, as the noise created by this chance inaccuracy will be reduced.

Schwartz & Metcalfe (1994) conducted a review of the studies to date which had analysed FOK accuracy using gamma. They identified 26 experiments which had used a forced choice recognition task, ranging from Yes-No recognition to two alternative forced choice up to 19 item forced choice. Despite a large variability in the materials used for these studies (general knowledge questions, picture identification, word pair learning, sentence completion), they found a modest but significant positive correlation between the number of test alternatives and the magnitude of gamma (r^2 = .21). Further examination of a subset of 14 studies which had followed a similar general knowledge question paradigm revealed a correlation of $r^2 =$ .63 between the number of alternatives and the magnitude of gamma. It therefore appears that the format of the criterion test can account for some of the variability in the value of gamma obtained across experiments. As with the previous issue of low stability, the effect of test alternatives is not an insurmountable problem for using gamma to assess FOK accuracy. Rather, it is a methodological quirk that should be taken into consideration when designing and interpreting the results of individual experiments and when examining results across experiments. Indeed, experiments in Chapters 2 and 3 used a four alternative forced choice recognition task to ameliorate the effect of the number of test alternatives.

The previous two issues can easily be accounted for by careful methodological design, for example by ensuring experimental manipulations are examined in terms of group effects rather than individual differences, using the same items if doing repeated tests, and interpreting experiments with different numbers of test alternatives with caution. However a third issue with gamma relates to instances of missing data and cannot be easily controlled for. As described above, gamma ranges from -1 to +1 and there are no restrictions on the values it can achieve. Thus it is possible, theoretically, for participants to show perfect relative accuracy (or indeed perfect relative inaccuracy). However, with real experimental data these values are only achieved when one of the elements of the gamma equation is equal to zero. If there are no false alarms or misses, the resulting gamma score will be 1. If there are no hits or no correct rejections, the resulting gamma score will be -1. This can potentially lead to instances where a participant receives a perfect positive gamma accuracy score despite making a large number of errors, or a perfect negative score despite making very few errors. Missing data points mean that insufficient information is available to ascertain the true sensitivity of the individual[6]. Ideally, the experimenter would continue testing until data had been obtained for each element of the gamma equation. This is not however possible practically, especially in the instance of episodic FOK tasks where a limited number of items are presented for encoding. Missing data becomes even more of an issue when two values are missing. In this instance, gamma is undefined, as too much information is missing to estimate the relative accuracy between the two variables. This is of particular concern when working with patient groups who may show impairments in metacognition. Here, participants may favour one FOK response over another, leading to a higher rate of missing data and undefined accuracy measures.

Souchay, Isingrini, & Gil (2002) encountered this difficulty when examining episodic FOK accuracy in Alzheimer’s Disease. However, rather than only using participants

---

[6] The Hamman coefficient, in contrast, is not affected by missing data due to the additive nature of the measure.

who gave data in all four response categories, Souchay et al. (2002) applied a correction advocated by Snodgrass & Corwin (1988) for old/new recognition data, which will be explained in detail below.

4.1.3 Recognition memory and signal detection theory

To fully understand where the correction originates from, we must take a detour into the recognition memory literature. Assessment of recognition data can form the same matrix as that of the FOK, shown in Table 4.1. In the FOK, the two dimensions making up the table are the FOK judgement (either a Yes or a No), and the recognition accuracy (either correct or incorrect). Within an old/new recognition task, participants are presented with a stimulus list for encoding, the 'old' item list. At test, these items are interspersed with items from a 'new' list, stimuli the participants have not previously encountered in the task. For each stimulus the participant must decide whether to respond 'yes' that the item is old and was encoded, or 'no' that the item is new. As with the FOK, there are four possible combinations of the two dimensions of recognition response and stimulus type, as shown in Table 4.2.

Table 4.2: Stimulus-response matrix for recognition memory test

	Recognition Response		
Stimulus	**Yes**	**No**	**Σ**
Old	a	b	$a + b$
New	c	d	$c + d$
Σ	$a + c$	$b + d$	$n = a + b + c + d$

To assess the accuracy of recognition, it is necessary to consider not only the hit rate but also the false alarm rate. Consider two participants who are presented with 20 items at test, 10 of which are old and 10 of which are new. Participant A responds 'old' to most items, thus receiving a hit rate of 9 and a false alarm rate of 8. Participant B also has a hit rate of 9, but a false alarm rate of only 2. If only the hit rate was used to assess accuracy, both participants would be deemed to have high recognition accuracy within the current task. However, it is clear that both participants are not equal in their ability to pick out old items. Thus the false alarm rate also needs to be considered, where the participant has judged an item to be 'old' when in fact they have not previously encountered it. It is the ability to discriminate between new items and old items which is the true assessment of accuracy in this task. The simplest way to measure this would be to subtract the false alarm rate from the hit rate to give a recognition score. In the present example, Participant A would thus have an accuracy score of 1, while Participant B would show much higher accuracy with a score of 7.

A more complex method of assessment is offered by signal detection theory (SDT: Green & Swets, 1966). SDT assumes that the individual is an active decision-maker who is trying to detect the signal (old or target items) from the noise (new or lure items) on each recognition trial. The simplest and most commonly used model for recognition memory is the one dimensional equal variance model (see Dobbins & Han, 2007). Each decision as to whether an item is old or new relies on one continuous, normally distributed attribute called familiarity. Target items have one probability distribution of familiarity, while lure items have a separate probability distribution of familiarity (see Figure 4.1). As old items tend to be more familiar than lure items, the target distribution sits to the right of the lure distribution. A decision criterion shows the response bias of the participant; above the criterion items are classed as old and receive a 'Yes' response, below the criterion items are classed as new and receive a 'No' response.

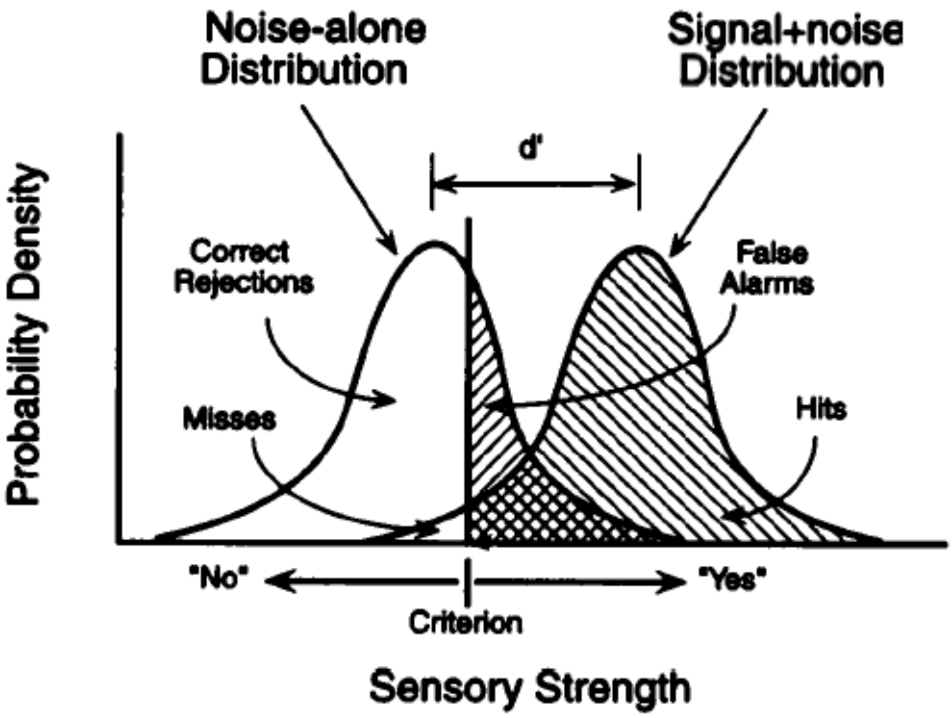


Figure 4.1: The equal variance signal detection model for recognition memory

Irrespective of the response bias, the *sensitivity* of the individual to the target distribution can be ascertained by the measure *d′*, which gives the distance between the two distributions. A large *d′* value implies that the target and lure distributions show little overlap, and the individual is able to easily discern which distribution the item in question belongs to. A small *d′* value shows that the individual has greater difficulty identifying the correct distribution. An approximation of *d′* is obtained from the hit rate and the false alarm rate by transforming them to a *z* score as follows (Macmillan & Creelman, 2005):

$$\boldsymbol{d' = z(H) - z(F)}$$

(3)

To refer back to our example participants, Participant A has a hit rate of 0.9 and a false alarm rate of 0.8, giving a *d′* of 0.36. Participant B also has a hit rate of 0.9, and a false alarm rate of 0.2, giving a *d′* of 2.93, again demonstrating the greater ability of Participant B to identify target items from lures compared to Participant A.

However, a problem arises when the hit rate is 1 or the false alarm rate is 0. Here, the measure is undefined and thus the ability of the participant to detect the target items from the lure items cannot be calculated. To avoid this issue, Snodgrass & Corwin (1988) recommended a correction (see Table 4.3). With this correction extreme values are impossible, thus *d'* can always be calculated. Although other corrections have also been applied to recognition data, the method above has been shown to be the least biased transformation of the data available (Hautus, 1995).

Table 4.3: Application of Snodgrass and Corwin's (1988) correction to recognition matrix

	Recognition Response	
Stimulus	**Yes**	**No**
Old	$(a + 0.5)/(a + b + 1)$	$(b + 0.5)/(a + b + 1)$
New	$(c + 0.5)/(c + d + 1)$	$(d + 0.5)/(c + d + 1)$

Souchay et al. (2002) proposed applying the same transformation to FOK data to ensure that no values are zero, allowing calculation of the gamma statistic even when there are instances of missing data. When applied to data from a previous study (Souchay et al., 2000), the adjusted gamma statistic was shown to be highly correlated with the unmodified gamma statistic (r = .90, p < .001; N = 61). It is worth noting that, as the unmodified gamma will include participants whose accuracy values are -1 and +1, this correlation will underestimate the relationship between the two measures. As per the recommendation of Snodgrass & Corwin (1988), the adjustment was applied to all data, not just to individuals where the gamma statistic was undefined. Although using an adjustment does to some extent constrain the values gamma can take (perfect values of -1 or +1 are no longer possible), this constraint is of minimal impact when considering the potential amount of extra data that can be included in assessment of FOK accuracy and the effect of experimental manipulations by applying the transformation, hence its use throughout Chapters 2 and 3 of the current thesis. In addition, unlike the HC which is constrained by the marginal values and this varies with each individual, the constraint for the correction is identical for each participant, so does not have any unwanted effect on the accuracy measure.

4.1.4 Signal detection theory and FOK

Signal detection theory has itself been considered as a potential accuracy measure for the FOK and metacognitive data more generally. Nelson (1984) originally rejected SDT due to the necessary assumption of normality required for the underlying distribution of the FOK. At the time of his consideration of the multiple measures available there was insufficient evidence to warrant this assumption, leading him to favour gamma over SDT. Recent observations have however indicated that a normal distribution is credible for metacognitive data (e.g. Benjamin & Diaz, 2008), leading to the suggestion that SDT should be applied to the FOK paradigm (Masson & Rotello, 2009). The question then arises as to what the SDT model of FOK should be. Masson & Rotello (2009), in their consideration of the associated issues and merits of gamma and SDT, assumed two underlying distributions similar to those used in recognition memory. One distribution represents the target items, the second distribution represents lure items, and an FOK criterion determines whether items receive a Yes or No FOK response. Thus the FOK model is identical to the recognition model with the recognition criterion replaced by the FOK criterion. However, this is not appropriate for the FOK where only target items are of interest. The FOK is made prior to recognition, it is the correct or incorrect recognition of the target items in light of the FOK judgement that allows for subjective hits, misses, false alarms and correct rejections. Therefore to include lure items is redundant, and the Masson and Rotello model is not a suitable representation of the FOK.

One possibility is that there are two underlying distributions that relate to positive and negative FOK judgements, with a recognition criterion allowing for the four possible outcomes. Indeed, to some extent this assumption of separate Yes FOK and No FOK distributions is implicit in the application of the adjusted gamma suggested by Souchay et al., (2002). The idea of there being a feeling of not knowing in addition to a feeling of knowing has received some support (Glucksberg & McCloskey, 1981; Klin, Guzmán, & Levine, 1997; Kolers & Palef, 1976; Liu, Su, Xu, & Chan, 2007; Luo, 2003), and Cheng (2010) attempted to create a novel measure of FOK accuracy based on this assumption. However the primary focus of FOK research has assumed that FOK varies on a single dimension, with low values corresponding to a No FOK and high values corresponding to a Yes FOK. Therefore it would be more appropriate to assume this concept of FOK judgements when considering a SDT based model.

Thus far we have a target item distribution with a recognition criterion, and a FOK distribution for the same target items with an FOK criterion. If we also assume equal variance for the two distributions (the simplest SDT model), there would be complete overlap of the two variables. However, as can be seen in Figure 4.2 (panels a and b), the four observable outcomes of hits, misses, false alarms and correct rejections are not possible in this model; only a single type of error can be made. One way to allow for both error types would be to rotate the FOK distribution around the point [0,0], as in panel c of Figure 4.2. The greater the rotation, the greater the errors made. However, rotation of the FOK distribution in isolation restricts the values the observations can be. Only two of the four observations can be truly independent. It is only by combining the placement of the FOK criterion and the angle of rotation of the FOK distribution that three of the possible four observations can vary. Thus the angle of rotation reflects FOK sensitivity, while the placement of the criterion reflects the bias in FOK responding.

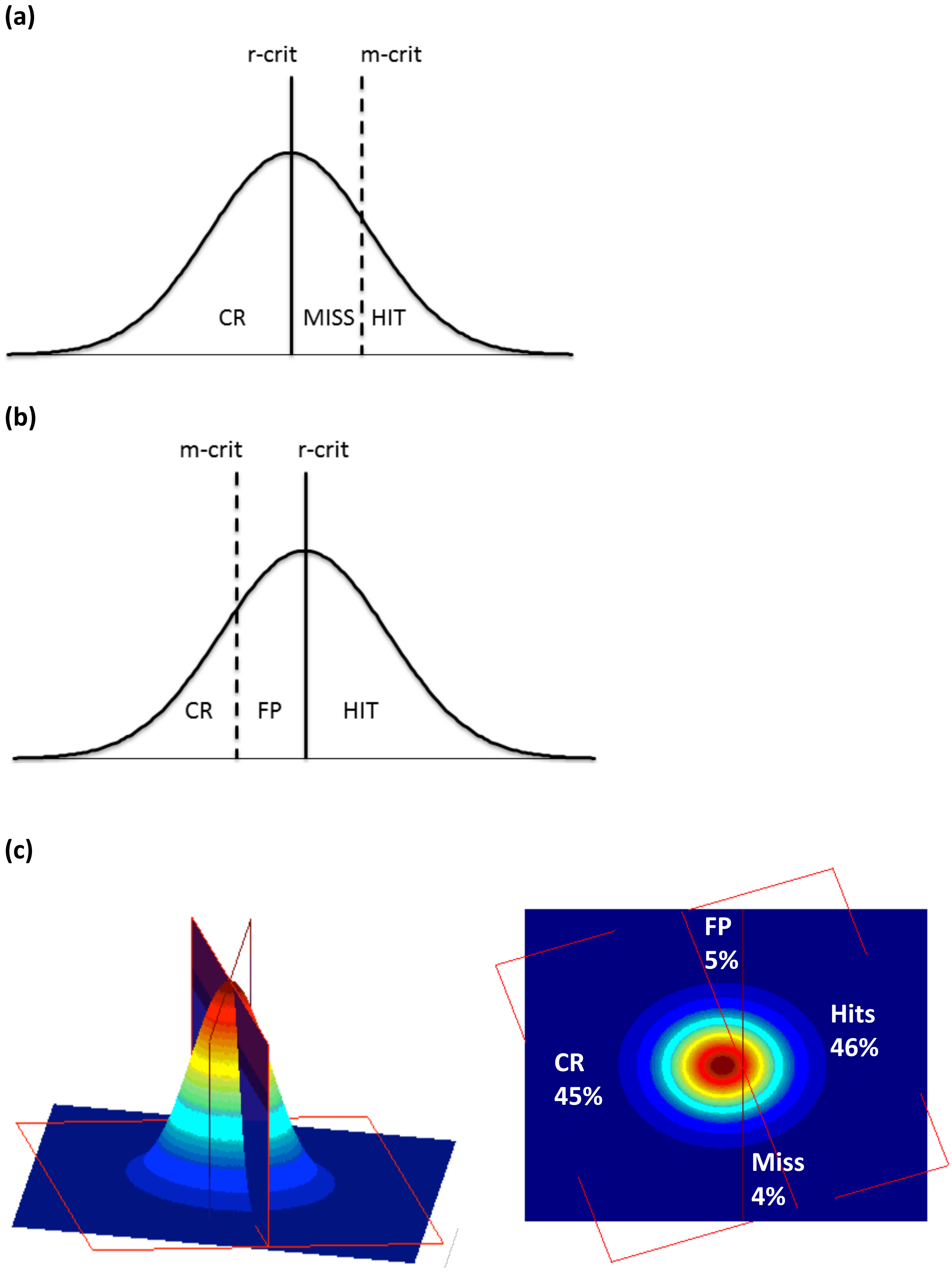


Figure 4.2: Development of the SDT model for FOK. Panels (a) and (b) demonstrate the problem with the standard SDT model. For conservative metacognitive judgements (a), m-crit falls above r-crit, resulting in the three observations of Hits, Misses, and CR (correct rejections). For liberal judgements, m-crit falls below r-crit, resulting in the three observations of Hits, Misses and FPs (false positives or false alarms). It is only by rotating the metacognitive distribution that all 4 observations can be made (c).

4.1.5 The model

In sum, I have proposed a two dimensional equal variance signal detection model from which nonrecalled items are drawn. The recognition distribution is assumed to have a mean of 0 and standard deviation of 1. A recognition criterion, named *r-crit*, determines the recognition response, with items falling above the criterion receiving a Yes recognition response and items below the criterion receiving a No. The placement of the criterion is thus determined by the summed proportion of hits and misses. The FOK distribution is also assumed to have a mean of 0 and standard deviation of 1. The FOK distribution is modified by the angle of rotation around the point [0,0], the parameter *theta*. A final parameter, *m*, quantifies the displacement of the metacognitive criterion from the recognition criterion. As with r-crit, values above the criterion result in a Yes FOK response and values below the criterion result in a No FOK response. We therefore have a total of three parameters; r-crit, *m* and *theta*.

The analysis of the suitability of the model for FOK will focus on the two metacognitive parameters *m* and *theta*. As with standard measures of metacognitive accuracy, the level of recognition is assumed not to affect FOK resolution (Nelson, 1984). *Theta* is expressed as a radian within the model, where 0 is equivalent to a rotation of 0° and *pi* is equivalent to a rotation of 180°. In order to allow comparison of the *theta* parameter to that of the gamma statistic, the cosine of *theta* is calculated. As with gamma, cos(*theta*) ranges from -1 (a rotation of 180°, an inverse relationship between judgement and recognition) to +1 (a rotation of 0°, perfect relationship between judgement and recognition).

The second parameter, *m*, is the distance between r-crit and m-crit. Positive values of *m* indicate that m-crit falls above r-crit, and that the individual is conservative in their metacognitive judgements. A negative *m* value indicates that m-crit falls below r-crit, and the individual is liberal in their metacognitive judgements. As already described, changes in *theta* or *m* in isolation would only allow two of the possible four observations to vary independently; it is only by allowing *theta* and *m* to interact with each other that three of the four observations can vary. Given that without *m* the proportion of hits-to-misses and correct rejections-to-false alarms would be linked, changes in *m* should track these relative proportions against each other. An error bias was thus calculated to quantify the proportion of old errors to the proportion of new errors based on the observed data, quantifying how likely people are to give Yes or No FOKs:

$$\boldsymbol{error\ bias} = \ln\left(\frac{\boldsymbol{miss}}{\boldsymbol{hit}}\Big/\frac{\boldsymbol{false\ alarms}}{\boldsymbol{correct\ rejections}}\right)$$

(4)

A positive error bias indicates conservative FOK judgements, a negative error bias indicates liberal FOK judgements. Thus if m is a measure of bias, it should reflect the changes in error bias based on the observed data.

4.1.6 Model fit with simulated data

In order to test the model, data were simulated for two levels of monitoring accuracy; chance vs. accurate. For the chance condition, 100 random numbers were generated in Excel using the *rand* function to represent 100 'item' responses for a total of 50 'participants'. The *rand* function returns an evenly distributed number between 0 and 1. Values greater than or equal to 0.75 were assigned as hits, values less than 0.75 but greater than or equal to 0.5 were assigned as misses, values less than 0.5 but greater than or equal to 0.25 were assigned as false alarms, and values less than 0.25 were assigned as correct rejections. For the accurate condition, the same procedure was used to assign random response to 40 of the 100 'items'. The remaining 60 'items' were automatically assigned as hits. Again, 50 'participants' were sampled. The mean percentage outcomes for each condition can be found in Table 4.4.

Table 4.4: Summary of percentage outcomes for simulated data in Chance and Accurate conditions

	Hits	Misses	False alarms	Correct rejections
Chance	24.82	24.16	25.5	25.52
Accurate	69.46	10.1	9.76	10.68

For both the chance and accurate condition, standard gamma scores and the error bias were calculated for each of the 50 participants. The proportions of hits, misses, false alarms and correct rejections were then entered into the model, and a solver function searched through possible parameters using least squares fitting to identify the optimal parameters. If the model is consistent with existing measures, we would expect strong positive correlations between cos(*theta*) and gamma, and between *m* and the error bias.

The parameters retrieved by the model and the equivalent statistics can be seen in Table 4.5. The manipulation of accuracy was successful, with gamma correlations higher for the accurate group compared to the chance group, $t(98) = 22.468$, $p < .001$. The model parameter of accuracy, cos(*theta*), was also significantly higher for the accurate group than the chance group, $t(98) = 18.944$, $p < .001$. In addition, one sample t-tests against zero were nonsignificant for the chance group (gamma $t(49) = 0.433$, $p = .667$; cos(*theta*) $t(49) = 0.561$, $p = .578$), whereas the accurate group were significantly above zero (gamma $t(49) = 35.487$, $p < .001$; cos(*theta*) $t(49) = 24.667$, $p < .001$). Thus the model parameter would appear to detect the presence of accuracy in a similar way to the gamma correlation.

Table 4.5: Retrieved model parameters and associated statistics for chance and accurate simulated data

		Parameters		**Associated statistics**	
		cos(*theta*)	***m***	**gamma**	**error bias**
Chance	***M***	0.011	0.099	0.011	0.021
	SD	0.136	0.129	0.177	0.398
Accurate	***M***	0.602	0.355	0.744	1.854
	SD	0.172	0.258	0.148	0.634

Further support for the correspondence of the two measures can be seen in Figure 4.3. For both the chance condition and the accurate condition, a clear relationship is shown between gamma and cos(*theta*). When examined using Pearsons' correlation coefficient, strong positive correlations are shown between the two measures in both the Chance, $r = .980$, $p < .001$, and the Accurate, $r = 0.924$, $p < .001$, conditions. These high correlations indicate that the two measures are indeed assessing the same metacognitive construct.

The second parameter, *m*, is a measure of bias. Positive values indicate conservative judgements and negative values indicate a liberal bias, as do values calculated with the error bias given in Equation 4. Both the Chance and Accurate conditions show a conservative bias with both measures. As already described, for the model to be deemed successful, *m* should map on to the error bias of the observations. A strong correlation between *m* and error bias is present within the Chance, $r = 0.933$, $p < .001$, and Accurate, $r = .911$, $p < .001$, conditions (see Figure 4.4). The model parameter does track the relative proportions of error as anticipated.

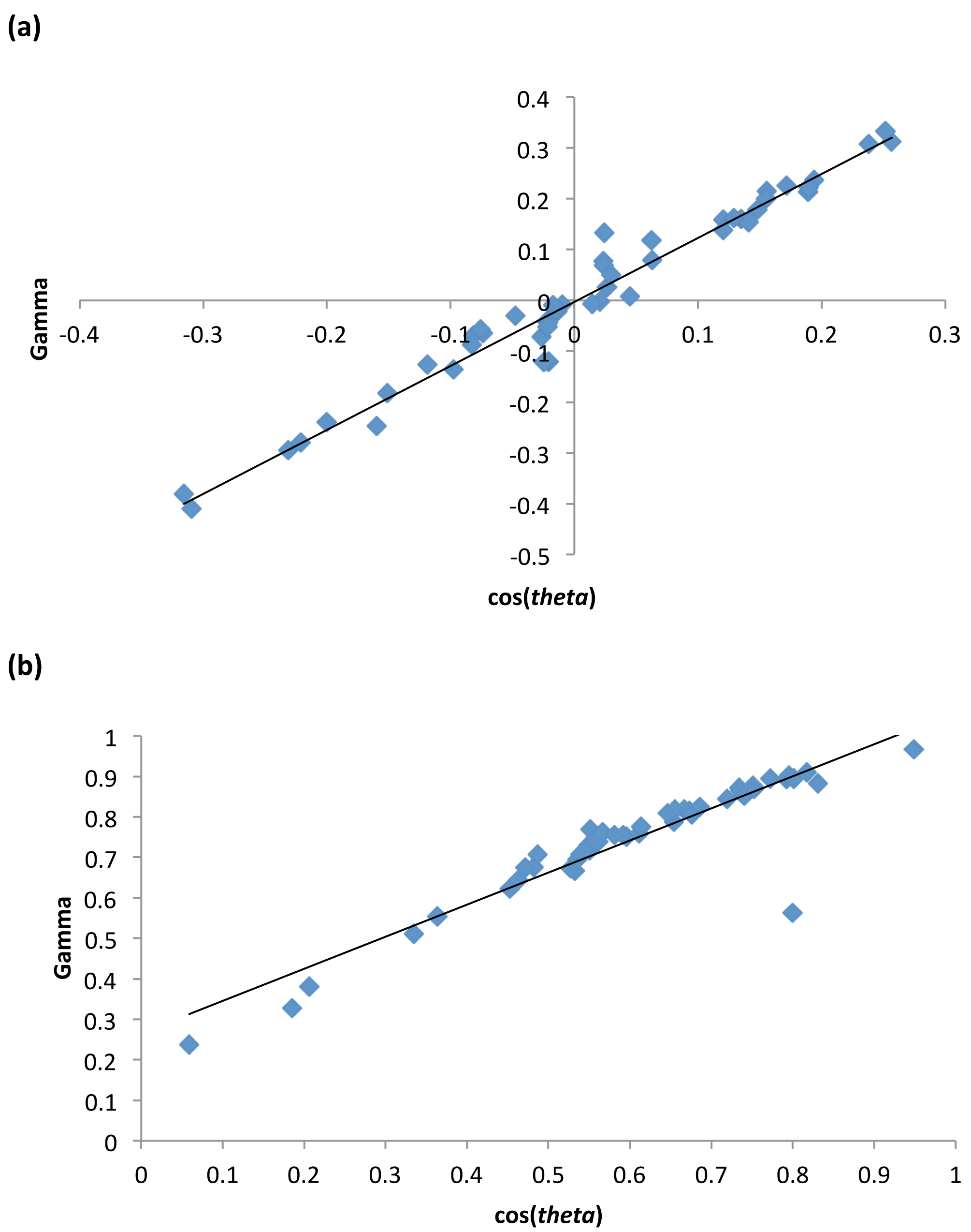
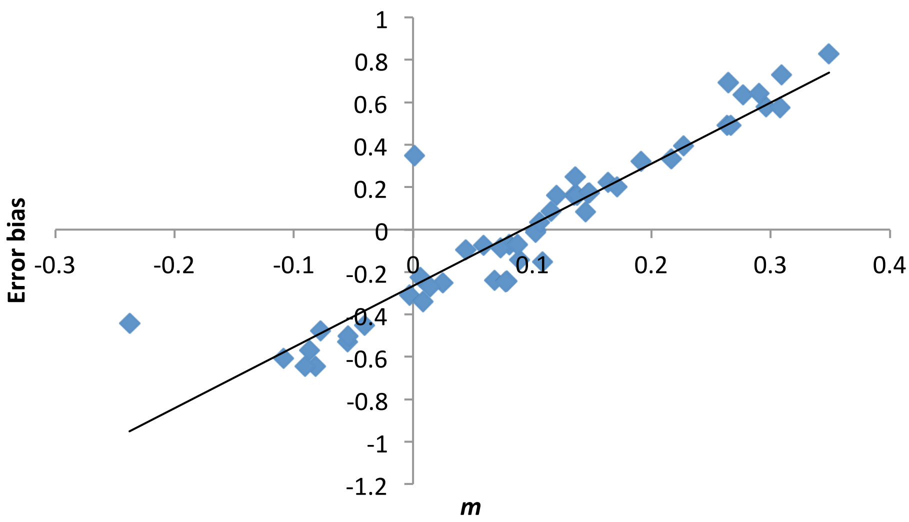


Figure 4.3: Scatterplot of gamma and model parameter cos(*theta*) measures of accuracy for (a) Chance condition and (b) Accurate condition

(a)



(b)

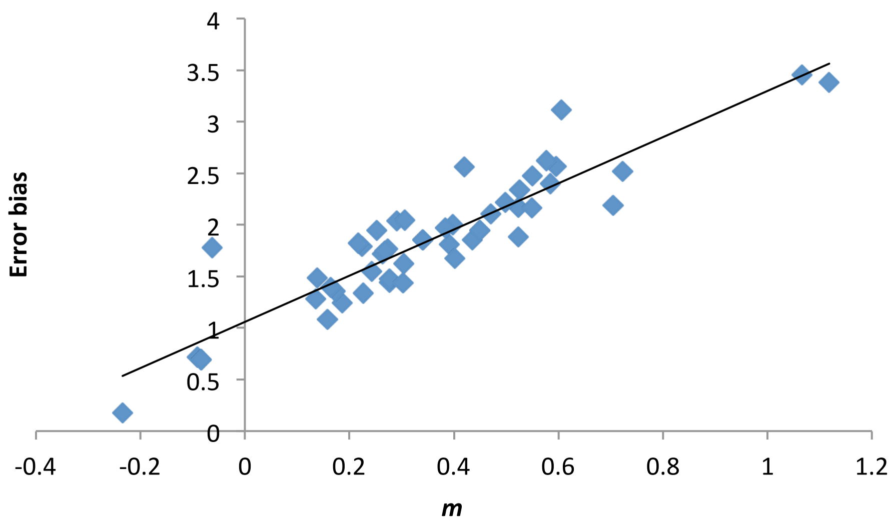


Figure 4.4: Scatterplot of model parameter m and error bias for (a) Chance condition and (b) Accurate condition

4.1.7 Model fit with experimental data

A demonstration of the applicability of the model to experimental data rather than simulated data is necessary to establish whether the model is a feasible approximation of the FOK. Experiments 2.1 and 2.2 in Chapter 2 both involved a Yes/No FOK response. Although the recognition task was a four alternative forced choice, as already noted the lure items are not of interest. Thus the recognition test is essentially a Yes/No recognition task for the target items only. We therefore have two sets of experimental data that can be examined with the model. To recap, Experiments 2.1 and 2.2 involved a semantic trial prior to episodic trials on the same materials, either general knowledge questions (Experiment 2.1) or French language learning (Experiment 2.2). The episodic data is of primary interest for the present test. We therefore used the data when filtered with the strict criterion, including only items which were not correctly recognised in the semantic trial prior to episodic learning. The model also encounters problems if observations are missing, leading us to only include data from those participants who reported all four observations. Due to the unresolved issue of whether older adults show a selective impairment in episodic FOK accuracy, we focused on the young adults only for this initial test of the model to minimise the influence of variables not controlled for in the model.

For Experiment 2.1 (the general knowledge quiz), 22 out of the 35 young adults who participated provided observations in all four possible outcomes. The proportions of hits, misses, false alarms and correct rejections for each participant were entered in to the model, and the parameters retrieved (see Table 4.6). As can be seen in Figure 4.5, a clear relationship is found between the traditional gamma measure and the cos(*theta*) measure. Indeed, the two measures show a significant large positive correlation, $r = .721$, $p < .001$. The correspondence between the two measures adds support to the previous finding that they are assessing the same metacognitive construct, and the rotation parameter *theta* can therefore be seen as another way of examining metacognitive accuracy.

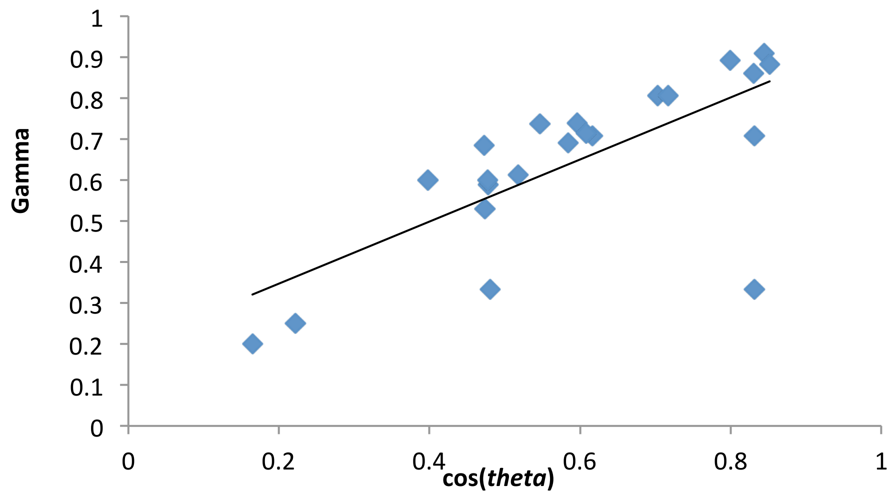
Table 4.6: Retrieved model parameters and associated statistics for data from Experiment 2.1 and Experiment 2.2

		Parameters		Associated statistics	
		cos(*theta*)	***m***	**gamma**	**error bias**
Experiment 2.1	***M***	0.593	0.035	0.645	0.120
	SD	0.195	0.426	0.206	1.500
Experiment 2.2	***M***	0.301	-0.333	0.228	-1.077
	SD	0.363	0.574	0.347	1.689

The second parameter, *m*, is small but positive, indicating participants tend to be slightly conservative in their FOK responding. That is, they are more likely to give a No FOK response than a Yes FOK response. As can be seen in Figure 4.6, the two variables are indeed closely related, with a strong positive correlation, $r = .876$, $p < .001$.

For Experiment 2.2 (the French study), of the 29 young adults who participated 21 gave complete sets of data. The parameters retrieved from the model can be found in Table 4.6. As previously, a strong positive correlation is found between gamma and cos(*theta*), $r = .739$, $p < .001$ (Figure 4.5). The value of *m* is negative, indicating a more liberal metacognitive criterion, unlike that found in Experiment 2.1. Again, a strong positive correlation was evident between *m* and the error bias, $r = .950$, $p < .001$ (Figure 4.6).

(a)



(b)

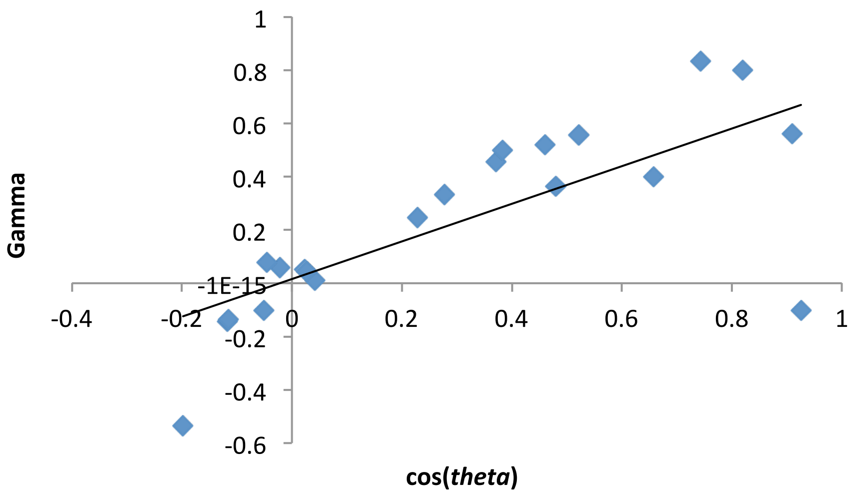


Figure 4.5: Scatterplot of gamma and model parameter cos(*theta*) measures of accuracy for (a) Experiment 2.1 and (b) Experiment 2.2

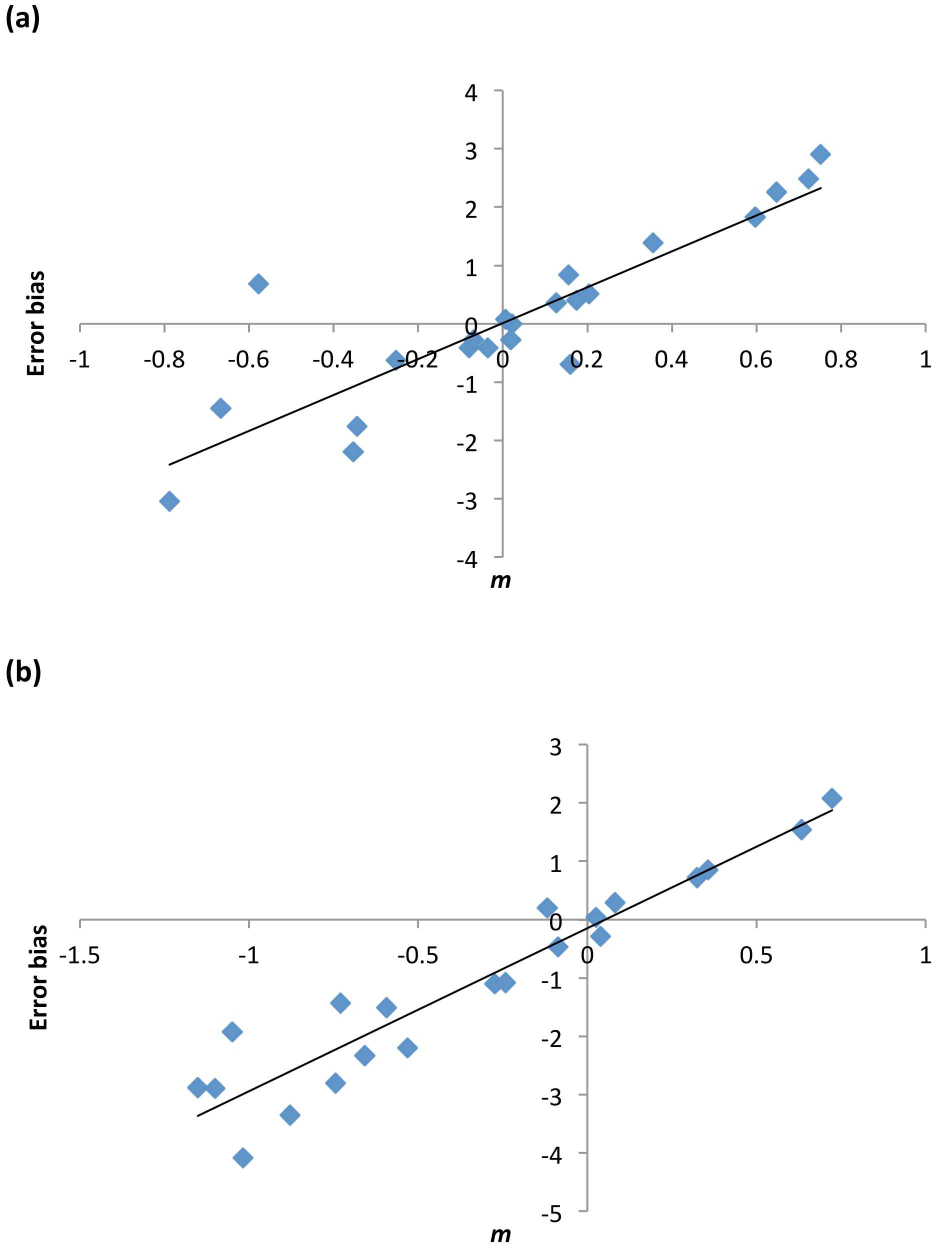


Figure 4.6: Scatterplot of model parameter *m* and error bias for (a) Experiment 2.1 and (b) Experiment 2.2

4.2 General discussion

The aim of the current chapter was to develop a model of FOK within the signal detection theory framework. Unlike previous models of FOK which relied on a Yes FOK distribution and a No FOK distribution, the present model assumed a single FOK distribution which was rotated from the recognition memory distribution. A metacognitive criterion also allowed examination of the bias in responding. The model was shown to fit both simulated and experimental data. The accuracy parameter was highly correlated with gamma, indicating the same metacognitive concept was being assessed, and the bias parameter also correlated well with error biases in the data. The model therefore not only provides a novel measure of accuracy which is consistent with the gamma measure of accuracy, it also goes beyond gamma to offer a second psychologically important parameter, an assessment of bias.

In addition to being mathematically compatible with the existing measure of FOK accuracy, the model is also compatible with the psychological theory of the FOK. FOKs are thought to be based on various mnemonic cues including familiarity with the cue itself and partial retrieval of information associated with the target (Koriat & Levy-Sadot, 2001). As per the metacognitive framework (Nelson & Narens, 1990, 1994), these mnemonic cues are used to create a model of the object level (the unrecalled item) at the meta level. Thus the FOK is based on a representation of the unrecalled item. When more diagnostic cues are used, this representation more closely reflects the object level, leading to greater accuracy in judgements. The SDT model presented here can also be considered in this way. The FOK distribution is a representation of the unrecalled item distribution. As more diagnostic information is available, the FOK distribution can closer reflect the unrecalled item distribution, thus reducing the angle of rotation between the two.

Signal detection theory has been extremely influential within recognition memory, driving much discussion about the underlying memory processes (e.g. Wixted, 2007; Yonelinas & Parks, 2007). So far, the underlying distribution (or indeed, distributions) of the FOK has received little attention, and application of SDT to the

FOK would enable this to take a much greater role in metacognitive theory. The current model assumes the FOK lies on a single distribution, with high levels corresponding to Yes FOKs and low levels corresponding to No FOKs. To simplify the model, this FOK distribution was assumed to be of equal variance to the recognition distribution, and to allow for independent observations a rotation of the FOK distribution from that of recognition was built in to the model. As a prototype model, the application of SDT principles to the FOK has been successful. This raises the possibility of more complex models being developed in the future. For example, many theorists have demonstrated that recognition memory is best modelled with an unequal variance model (e.g. Heathcote, 2003; Ratcliff, Sheu, & Gronlund, 1992). It may also be that an unequal variance model of FOK may be more appropriate. Alternatively, it may be that the FOK consists of two distributions, a feeling of knowing and a feeling of not knowing, as suggested by a small number of researchers (Glucksberg & McCloskey, 1981; Klin et al., 1997; Liu et al., 2007; Luo, 2003). If this is indeed the case, then FOK may be modelled in a more similar way to that of recognition memory. The bias measure also provides a novel way to explore the FOK. For example, older adults have typically been found to be underconfident when making metacognitive judgements (e.g. Marquié, Jourdan-Boddaert, & Huet, 2002). The current model would allow further examination of this effect within the FOK, how it may interact with accuracy, and whether it is possible to adjust bias in order to improve metacognitive performance. These considerations provide novel avenues to explore both theoretically and empirically.

Within the current chapter the focus has been on the simplest form of FOK, that originally proposed by Hart (1965) consisting of a 2 x 2 design. If we assume FOK is normally distributed as SDT requires rather than an absolute state, the 2 x 2 paradigm is the most imprecise method of assessing the FOK. Both gamma and SDT have problems in the 2 x 2 case due to tied observations. Tied observations occur either when a pair of observations are from stimuli belonging to the same class or evoke the same response from a participant. Goodman & Kruskal (1954) excluded these ties from the gamma calculation, however this reduces the number of observations on which gamma is calculated, thus reducing its stability (Spellman, Bloomfield, & Bjork, 2008). Although some have suggested an adjustment be applied when ties do occur (e.g. Kim, 1971; Somers, 1962), Gonzalez & Nelson (1996) argued that this was wholly dependent on the nature of the ties. With limited options on how to classify a response, for example as a Yes FOK or a No FOK, ties are ambiguous. Two items which receive the same response could either truly be judged as being the same, or they could be slightly different yet with only two options are assigned the same category. Gonzalez and Nelson recommend that these ambiguous ties not be included in the calculation of gamma.

The use of a rating scale allows finer judgements to be made and reduces the number of ambiguous ties (Masson & Rotello, 2009). Gamma easily transfers to a 2 x n design, thus can still be used as a measure of FOK accuracy. Reducing ambiguous ties will subsequently allow more observations to be used in the calculation of gamma, improving its stability. Ratings scales also provide greater resolution for SDT, again allowing for more complexity to be incorporated into the model proposed in the current chapter. However, one consideration with ratings scales is how fine a scale to use. The more available categories, the closer the model will be able to fit the available data. However, the model fails when data is missing, as does the traditional gamma correlation. With episodic FOKs only a limited number of observations that can be obtained, thus there is no guarantee that participants will provide observations in all response categories. Some researchers have used the approach of obtaining ratings from participants on a scale from chance to 100% with the experimenter assigning responses to bins to allow analysis (e.g. MacLaverty & Hertzog, 2009).

This post-hoc assignment to response categories was used in Experiment 3.1 and 3.2 in the current thesis. For analysis within Chapter 3, responses were assigned to one of three categories. It would also have been possible to assign the data to five different categories. Would this have affected the conclusions made? As can be seen from Table 4.7, the gamma correlations achieved do show some variation depending on the number of categories used, with five categories giving higher estimates of accuracy than three categories. The impact of the response categories used, whether decided prior to testing or after testing, is thus of great importance and should be a high priority for further exploration if the use of SDT is to continue within the FOK literature.

Table 4.7: Comparison of Gamma correlations for FOK data with post-hoc splits into 3 response categories and into 5 response categories

	FOK Before	FOK After	Control
3 category gamma	0.269 (0.62)	0.221 (0.48)	0.319 (0.44)
5 category gamma	0.285 (0.54)	0.251 (0.41)	0.360 (0.41)

In sum, the current chapter has proposed and supported an equal variance model of FOK. The model provides a measure of accuracy which appears to measure the same underlying construct as gamma, the predominant existing measure of FOK accuracy. The model also suggests an additional measure of bias which aids conceptualisation of the FOK. Further development of the model is warranted and required as currently it cannot reliably calculate accuracy and bias when data is missing, a limitation shared with the gamma statistic. This leads to the model becoming quickly overparameterised. In addition, the model needs to be mathematically formalised as it is currently computationally expensive to fit. This will assist in creating more complex models as well as providing possible solutions as to how to deal with missing observations. Empirical tests can then be designed to test the model more rigorously, for example by experimentally dissociating *theta* and *m*.

Whether or not SDT can or will replace gamma as the primary measure of FOK accuracy over the next few years is a difficult one to answer. Although there are a number of potential issues with the gamma statistic, most of these can be avoided with careful experimental design. For example, the influence of the number of test alternatives identified by Schwartz & Metcalfe (1994) can be controlled for by ensuring the same number of lure items are used when comparing across studies. A way to account for missing data has also been proposed by Souchay et al. (2002). SDT provides a promising new way to conceptualise the FOK, and by refining its application and developing new models it will encourage advances in the theories of how FOK are formed and the underlying processes involved. Until then, the gamma remains the best available measure, despite its flaws. Spellman et al. (2008, p. 112) quoted Winston Churchill's view of democracy to encapsulate the adoption of gamma within the metacognitive literature: "Democracy," said Sir Winston, "is the worst form of government except all those other forms that have been tried from time to time."

Chapter 5. Global Metacognitive Judgements at Retrieval Failure

5.1 General introduction

While a number of metacognitive judgements have been administered as both item-by-item and global measures, for example judgements of confidence (JOCs) and JOLs, the FOK has thus far only been used as an item-by-item judgement. Item-by-item judgements are typically used within the experimental cognitive psychology literature when examining normal healthy adults (e.g. Koriat & Ma'ayan, 2005; Nelson & Dunlosky, 1991), whereas global judgements have been primarily used in the neuropsychological literature (e.g. Moulin, Perfect, & Jones, 2000) and for studies into metacognitive development in children (e.g Flavell, Friedrichs, & Hoyt, 1970). The addition of a global FOK measure, or indeed any global measure of retrieval failure, would allow further assessment of metacognition in these special populations.

Evidence from JOCs and JOLs suggests that the differences between item-by-item and global judgements are more than simply methodological and their applicability to different populations, however. The logical assumption would be that global judgements reflect the summation of each item-by-item judgement, thus both measures would produce the same overall level of judgement and accuracy. This is not, however, evident in the data. JOCs show participants to be overconfident in their performance when asked to give item-by-item judgements, whereas global judgements lead to a lower level of overconfidence or may even show underconfidence (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Griffin & Buehler, 1999; Griffin & Tversky, 1992). A similar pattern is also shown for JOLs, with item-by-item JOLs showing overconfidence and global JOLs showing less overconfidence (Liberman, 2004; Mazzoni & Nelson, 1995). This has been termed the aggregation effect (Griffin & Tversky, 1992), and is proposed to occur due to differences in the referent used. For item-by-item paradigms, the judgement is made on the characteristics of the particular item in question. For global paradigms, the judgement is based on the characteristics of the individual's prior success at retrieving items under similar conditions. This is somewhat similar to the experience and theory based judgements examined by Koriat, Bjork, Sheffer, & Bar (2004). Item-by-item judgements could be conceptualised as experience based, relying on the mnemonic cues available at the time of the judgement. In contrast, global judgements are more theory based, depending on knowledge about how memory works generally and knowledge of how memory has performed within similar circumstances.

Recently, there has been a diversification in the types if metacognitive judgements which have been examined experimentally. Traditionally, researchers have focused on the types of judgements identified by Nelson & Narens (1990, 1994) when they presented their framework, such as JOCs, JOLs and FOKs. And yet there are numerous subjective experiences associated with memory, and cognition more generally, which may be available to study. Tauber & Rhodes (2012) proposed the judgement of retention (JOR), a prediction of how long an item would be remembered for. This assessment of monitoring is similar to a JOL, in that it is a judgement of future performance made immediately after learning. Rather than being a percentage, the JOR is a prediction of the number of minutes an item will be remembered for, thus providing a different scale (i.e. time) on which to assess memory awareness. Indeed, over a series of six experiments the authors noted that JORs provided more realistic predictions of future memory performance than did JOLs.

Although a number of measures have been proposed that evaluate the ability to predict future memory performance (Kelemen, 2000; McCabe & Soderstrom, 2011; McGillivray & Castel, 2011; Tiede & Leboe, 2009), metacognitive ability at memory retrieval has not developed as extensively. The key question when considering the development of a novel metacognitive measure is whether it relates to an existing memory experience and whether this experience drives memory behaviour. For example, after making a JOR participants selected a higher number of items to restudy than after making a JOL, thus influencing learning behaviour (Tauber &

Rhodes, 2012). It is therefore important to ask whether it is logical to obtain a global measure of the failure to retrieve a memory. Consider the following examples. When a student is revising a topic they are not memorising discrete items, as is typically assessed in an FOK task. They will be learning and revising a large amount of information and, to inform their study, they will be judging how much of this information is currently not available for explicit recall. This will then dictate memory related behaviours, such as deciding which areas to restudy and how much time and effort to devote to the topic. Similarly, when shopping in the supermarket it is common to experience the sensation that you are missing a certain number of items from your list even though you cannot state what those items are. This sensation will drive further attempts at recall, by thinking through recipes for planned meals or even visiting certain areas of the supermarket to cue the memory. The current FOK paradigm focuses on individual items, and assessments on whether that single fact could be recognised at future test. The metacognitive literature has not yet fully considered the stage before this, the assessment of the volume of nonretrievable information in memory which then determines memory behaviour.

In both of the above examples the rememberer is able to make a judgement about how much information has not been retrieved, even though the detail of the information is not available. The novel monitoring judgement proposed here is therefore termed a judgement of retrieval failure, or JORF. It is the ability to assess on a global level the information remaining in memory that is not currently accessible. This type of judgement circumvents one of the drawbacks of the standard FOK paradigm, namely that the participant knows a target item must exist for each cue item irrespective of their subjective experience. Thus the judgement is made on the assumption that for each cue a specific item has been encoded, and the memory sensation is evaluated from this basis. The JORF in contrast is a global judgement, and so multiple target items are associated with a single cue item. Therefore the participant does not receive any prompt as to whether a memory should exist for each specific target; instead they need to judge the number of target items themselves.

A further difference between the JORF and the FOK is the type of judgement that is made. Although it occurs at memory retrieval, the FOK is a predictive metacognitive judgement, like the JOL. It is an assessment of how likely an item will be accessible to memory at a *future* test. The JORF in contrast is a postdictive judgement, similar to the JOC. It is an assessment of memory performance which has already occurred. This important distinction therefore raises the possibility of a second novel metacognitive judgement to be assessed in conjunction with the JORF; a global FOK or GFOK. If people are able to assess the amount of information that they have failed to retrieve, it should also be possible to make a prediction about how much of this information will be retrievable at a later test. Note however that the magnitude of the global FOK will depend on the magnitude of the JORF; first an assessment must be made of how much information has not been recalled, then an assessment must be made of how much of the unrecalled information may be retrieved later. The two measures should therefore be closely related.

It is important to note that one study has examined judgements of information that has not been retrieved. Halamish, McGillivray, & Castel (2011) asked young and older adults to study items from five different categories for a later recall test. Items were presented in a single intermixed list, with varying numbers of exemplars from each category comprising the study list. Following free recall, participants were asked to judge the number of items from each category that they had not remembered, termed a monitoring of forgetting (MOF) judgement. They observed that the judgements of both young and older adults closely resembled the actual amount of information that had been ‘forgotten’. As expected, older adults recalled less, but they also judged that there were more items remaining than the younger adults, thus showing that they were aware of how much they had failed to retrieve.

The above study demonstrates that participants have some access to the information required to make accurate global judgements. The issue of whether this is a true judgement of forgetting is, however, open to debate. The term 'forgetting' is typically conceptualised as meaning that the item has been sufficiently encoded in the first place and was at some point available to explicit recall. Whether the items presented by Halamish et al. (2011) had, at some point, been recallable is unknown. Thus we propose that the judgement is more of a judgement of retrieval failure. This removes the implication that the item has been sufficiently encoded to allow a clear representation or memory trace. Clearly a level of encoding has been achieved by the participants in the Halamish et al. (2011) study as their judgements do reflect the number of items that remain. However the term 'retrieval failure' is less ambiguous regarding the memory status of the items in question.

As already discussed, the FOK and metacognitive judgements more generally rely on access to partial information about the target and the familiarity of the cue (Koriat & Levy-Sadot, 2001). When a JORF is made the judgement of the number of remaining items could be based on the distinctiveness of the partial information retrieved in addition to the volume of information. For example, if the words 'happy' and 'sad' had not been recalled, knowing the valence of each item would allow the rememberer to correctly assess there are two separate items in memory, and consequently give a JORF of 2. Alternatively, if the words 'happy' and 'joyful' had not been recalled, knowledge of the valence would not be diagnostic and could lead to an incorrect JORF of 1. This is similar to the distinctiveness heuristic to reduce false memory production (Dodson & Schacter, 2001; Dodson & Schacter, 2002).

It is important to consider the contribution of both episodic and semantic memory processes to the JORF, as this may influence the accuracy of the JORF given. A procedure similar to that of Halamish et al. (2011) would be classed as an episodic memory task, the combination of exemplars being unique to the time and place that they are encoded. Therefore episodic memory processes will be recruited. However, the use of the category name as the cue may also lead to semantic processes being involved, leading to members of that category being identified at recall despite not being present during encoding. This false memory due to semantic relatedness has been observed to happen with the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995). Here, a list of words is presented at encoding which are related to a non-presented lure item. At recall test, the lure is often reported in addition to the encoded items. This robust effect is thought to occur due to the associative nature of memory; the non-presented lure becomes activated due to a spread of activation from semantically related items (Gallo, Roediger, & McDermott, 2001). This level of activation then reaches a sufficient threshold for the item to be verbally reported. Thus within the JORF task semantically related items may receive a level of activation which could be confused with partial retrieval of items which were presented at encoding, leading to inaccurate assessments of the magnitude of retrieval failure. Additionally, a DRM-like effect could occur during recall, with items reported as being present when they were not.

In addition to this semantic spread of activation, false recall and inaccurate JORFs may also occur due to a combined failure of a generate-recognise strategy and source memory errors (Anderson & Bower, 1972; Johnson, Hashtroudi, & Lindsay, 1993). The generate-recognise strategy consists of two processes, an initial generation of information followed by a recognition evaluation of whether the item was presented at encoding. If the rememberer falsely attributes generation to a recent encounter, thus making a source memory error, they may either verbally report the item as being present or include the item within their JORF prediction.

A similar concept occurs in fuzzy trace theory (Brainerd & Reyna, 2002; Gernsbacher, 1985). According to fuzzy trace, two independent memory traces are formed during encoding: a gist-based trace and a verbatim trace. The gist trace represents the underlying meaning of the memory, so is related to the semantic aspects. The verbatim trace represents the exact characteristics of the memory, and so can be considered as the episodic aspects. Retrieval of the verbatim trace thus allows for two semantically related items to be identified as either being present at encoding or not. However, the verbatim trace decays at a faster rate than the gist trace (Gernsbacher, 1985), thus making it more difficult to discriminate between semantically related items, leading to errors of recall and potentially errors at JORF.

In sum, these three mechanisms of spread of activation, failure of source monitoring, and fuzzy trace theory in combination with the use of a semantically structured task suggest that the involvement of semantic processes will occur and may lead to errors of recall and also JORF within an episodic task. It is therefore the aim of the current chapter to not only establish the sensitivity of the JORF to classic memory manipulations, but also to examine the role of semantic and episodic memory and the effects of reducing episodic memory accuracy either experimentally or within a memory-impaired clinical population.

A core consideration of the JORF is its sensitivity to list length. As list length increases, so too will the number of items which are not recalled (Ebbinghaus, 1913). This should lead to appropriate increases in JORF to reflect the increase in retrieval failure. To date, the influence of list length on metacognitive judgements has received little attention. A single study by Tauber & Rhodes (2010) has considered the impact of the amount of material to be learned on JOLs. Here, JOLs were found to be sensitive to list length, increasing as list length increased, and yet judgements were poorly calibrated. As list length increased, participants became less accurate in their predictions of future performance. The list lengths used by Tauber and Rhodes (10 items, 60 items or 100 items) were much larger than would be practical within the current paradigm, nonetheless the findings do suggest that metacognitive judgements are influenced by the amount of material presented. It is therefore expected that JORFs will vary in a similar way to retrieval failure.

The current chapter presents three studies examining JORFs and GFOKs and their sensitivity to some fundamental characteristics of memory performance. Firstly, the effect of list length was investigated in Experiment 5.1 with a large sample of healthy undergraduate participants. In addition, Experiment 5.1 also demonstrated the effect of delay on the JORF and GFOK. Similar to the set size argument, as delay increases so too will retrieval failure, and therefore so too should JORF while GFOK decreases. However, as already discussed in relation to Experiments 3.1 and 3.2, the amount of diagnostic information on which these judgements may be based will also deteriorate over the delay, and so these judgements may become less accurate. Experiment 5.1 assessed whether this is indeed the case.

A fundamental distinction in memory is that of episodic and semantic memory. Semantic memory is formed over many exposures and relates to general information and knowledge about the world. In contrast episodic memory is linked to a specific time and place, the memory for a single, unique event. The type of memory can affect the magnitude and accuracy of metacognitive judgements, as already discussed in the context of FOKs and ageing in Chapter 2. Experiment 5.2 therefore aimed to assess the accuracy of JORFs for semantic and episodic memory separately, to establish whether this measure is also sensitive to the semantic-episodic distinction. The final experiment within this chapter, Experiment 5.3, examines the JORF and global FOK in a clinical population. Global judgements have proven to be useful in examining metacognitive abilities in patient populations. The utility of these judgements was established within a patient group with memory impairment. To achieve this, a sample of participants diagnosed with probable dementia was recruited, along with a healthy older adult control group.

5.2 Experiment 5.1: Judgements of retrieval failure and the global FOK

5.2.1 Method

5.2.1.1 Participants

Psychology undergraduate students from the University of Leeds were recruited for the study. The study formed part of a level three module to demonstrate subjective experiences in memory. Due to participants being tested over a long delay (14 weeks) not all participants were available for testing at both time points. Full data was provided by 48 participants in total.

5.2.1.2 Materials and procedure

Five categories of words were chosen to form the word lists: fruits, colours, creatures, French words and nonwords. Presentation order and list length were the same for all participants due to the study forming part of a lecture. Thus all participants were asked to encode 13 fruits (fruits_13), followed by seven French words (French_7), 10 colours (colours_10), 17 creatures (creatures_17) and nine nonwords (nonwords_9). List length was varied to measure whether judgements were sensitive to the size of each category. Each list was presented via Powerpoint for the equivalent of two seconds per item i.e. fruits were presented for 26 seconds, French words for 14 seconds etc. Items were randomly distributed across the Powerpoint slide, with the category name written in bold at the top and centre of the slide. Category names were presented in Calibri font size 24, with category items presented in Calibri font size 18, and were projected onto a screen measuring 212cm by 216cm. Recall was tested immediately after encoding. Participants were provided with an answer sheet listing the category names in the same order as presented during learning. They were asked to recall any items they could from each list and not to guess.

Following recall, two questions were asked to measure Judgement of Retrieval Failure (JORF) and Global Feeling of Knowing (GFOK) for each category. JORFs were elicited with the prompt `How many other items on the list were there that you haven't recalled above?'. GFOKs were elicited with the prompt `How many of these that you didn't recall will you recognise when they are shown again?'. Participants were free to give any number they chose. Upon completion of each judgement, answer sheets were collected and the recognition test was given. The recognition task comprised a single sheet with each category name listed followed by a number of category exemplars and a Yes/No response option. Each category was presented in the same order as at learning. The category exemplars consisted of the correct target items plus an equivalent number of lures i.e. 26 fruits, 14 French words etc. Participants were asked to indicate which words were presented as part of the list by circling the Yes response option.

A surprise second recall and recognition test was given 14 weeks later. Answer sheets were identical at this second test i.e. the same category order and the same distracters in the recognition test. JORF and GFOK measures were also obtained.

5.2.1.3 Analysis

Recognition accuracy was analysed using a discrimination index, i.e. hits minus false alarms. Due to differences in list length, the number of items correctly recalled, discrimination index, and JORF predictions were divided by category size to provide a proportion. For example, if Participant A recalls 3 correct colours out of 8 total, accuracy is 0.375. Whereas if Participant B recalls 3 correct creatures out of 6 total, accuracy is 0.5. To examine the accuracy of JORF predictions, judgements need to be compared to actual retrieval failure i.e. 1-proportion recall. Thus if Participant A gives a JORF of 5, proportion JORF is 0.625, and actual retrieval failure is also 0.625. Participant A is therefore accurate in their JORF. In contrast if Participant B gave a JORF of 1, proportion JORF would be 0.167, while actual retrieval failure is 0.5, thus reflecting an inaccurate JORF.

The GFOK measure is dependent on JORF assessment, therefore the GFOK measure is divided by the JORF to provide a proportion. This can then be compared to the proportion of actual recognition of unrecalled items. It is therefore possible for participants to give an inaccurate JORF but an accurate GFOK. For example, if a participant gives a JORF of 8 when there are actually 4 non-retrieved items, their JORF will not be accurate. However, if their raw GFOK is 6 of the 8 items judged to be remaining, and their raw recognition of unrecalled items is 3 of the actual 4 items unrecalled, then proportionally they are accurate (0.75 in both cases). Thus the participant shows relative metacognitive accuracy, in that they are able to accurately predict what proportion of unrecalled items they will recognise at a later test.

Intrusion errors at recall were not converted to proportions.

5.2.2 Results

Due to differences in list length, all five categories were analysed on an individual basis. To control for multiple comparisons, alpha was set at 0.01.

5.2.2.1 Memory

Immediate recall and recall following the 14 week delay were compared using paired samples t-tests for each category. As expected, a significant effect of delay was observed for all categories, with recall accuracy higher before the delay than after (French_7 $t(47) = 10.475$, $p < .001$; nonwords_9 $t(47) = 7.864$, $p < .001$; colours_10 $t(47) = 11.891$, $p < .001$; fruits_13 $t(47) = 15.918$, $p < .001$; creatures_17 $t(47) = 10.456$, $p < .001$;). Mean recall is 1 minus retrieval failure as presented in Figure 5.1. Likewise, the discrimination index measure of recognition shows a significant effect of delay as expected, with greater discrimination accuracy before the delay than after for all categories (see Table 5.1). In most cases this is due both to a decrease in hits over the delay and an increase in false alarms. However for nonwords, only a decrease in hits is found, with no effect of delay on false alarms (see Table 5.2).

Table 5.1: Discrimination index at immediate and delayed recognition for each category

	Immediate	Delayed
French words_7	0.676 (0.228)	0.298 (0.255) ***
Nonwords_9	0.456 (0.298)	0.111 (0.130) ***
Colours_10	0.665 (0.184)	0.323 (0.186) ***
Fruits_13	0.731 (0.186)	0.289 (0.196) ***
Creatures_17	0.599 (0.207)	0.153 (0.142) ***

Note: *** p < .001, paired sample t-tests, two tailed

Table 5.2: Hits and false alarm rates for each category at immediate and delayed test

	Hits		False alarms	
	Immediate	*Delayed*	*Immediate*	*Delayed*
French words_7	0.827 (0.138)	0.554 (0.218)***	0.152 (0.154)	0.256 (0.186)**
Nonwords_9	0.588 (0.217)	0.227 (0.193)***	0.132 (0.141)	0.116 (0.148)
Colours_10	0.844 (0.092)	0.646 (0.200)***	0.179 (0.134)	0.323 (0.172)***
Fruit_13	0.894 (0.074)	0.667 (0.150)***	0.164 (0.143)	0.378 (0.177)***
Creatures_17	0.750 (0.135)	0.463 (0.171)***	0.151 (0.128)	0.310 (0.179)***

*Note: ** p < .01, *** p < .001, paired samples t-tests, two tailed*

The pattern of intrusion errors is less consistent over the categories (Table 5.3). Colours_10, $t(47) = 4.924$, $p < .001$, fruits_13, $t(47) = 3.840$, $p < .001$, and creatures_17, $t(47) = 4.389$, $p < .001$ all show a significant increase in the number of intrusion errors over the delay. However, French words_7 show no such effect, $t(47) = 2.480$, $p = .017$, nor do nonwords_9, $t(47) = .0340$, $p = .736$.

Table 5.3: Intrusions errors at immediate and delayed recall for each category

	Immediate	Delayed
French words_7	0.104 (0.309)	0.313 (0.512)
Nonwords_9	0.146 (0.412)	0.104 (0.722)
Colours_10	0.438 (0.796)	1.750 (1.769)
Fruits_13	0.396 (0.140)	1.083 (1.145)
Creatures_17	0.229 (0.555)	1.042 (1.271)

5.2.2.2 JORF

The proportionalised data for JORF and actual retrieval failure were analysed using a 2 (delay: immediate and 14 weeks) x 2 (measure: JORF and actual retrieval failure) repeated measures ANOVA (see Figure 5.1). For French words_7, a main effect of delay was found, $F(1,47) = 67.561$, $p < .001$, $\eta_p^2 = .590$, as expected with both measures being higher after the delay than before. A main effect of measure was also found, $F(1,47) = 46.759$, $p < .001$, $\eta_p^2 = .499$. JORFs were higher than actual retrieval failure, suggesting participants overestimated the number of items that they were unable to recall, predicting that there were more items on the list than there actually were. A significant interaction between delay and measure was also present, $F(1,47) = 19.198$, $p < .001$, $\eta_p^2 = .290$. Paired samples t-tests indicate that JORFs are higher than retrieval failure at immediate test, $t(47) = 4.103$, $p < .001$, and after the 14 week delay, $t(47) = 6.486$, $p < .001$. In addition, both JORFs, $t(47) = 6.589$, $p < .001$, and retrieval failure, $t(47) = 10.475$, $p < .001$, increase over the delay. The interaction appears to be caused by greater overestimation of retrieval failure after the delay than prior to the delay. Thus after 14 weeks the JORF becomes less accurate. Nonwords_9 also show a main effect of delay, $F(1,47) = 32.663$, $p < .001$, $\eta_p^2 = .410$, with scores lower before the delay than after. However, no main effect of measure is shown, $F(1,47) = 1.540$, $p = .221$, $\eta_p^2 = .032$. JORFs and actual retrieval failure are equivalent, indicating participants are, overall, able to accurately predict how many items they have not been able to recall at test. No interaction was found, $F(1,47) = 4.561$, $p = .038$, $\eta_p^2 = .088$, indicating that the delay did not lead to greater overestimation. Participants were as accurate immediately as they were 14 weeks later..

Colours_10 show a similar pattern to French words_7, with a main effect of delay due to increased forgetting over time, $F(1,47) = 81.977$, $p < .001$, $\eta_p^2 = .636$, and a main effect of measure, $F(1,47) = 50.134$, $p < .001$, $\eta_p^2 = .516$, with JORFs being higher than actual forgetting. However, no interaction was present, $F<1$, indicating participants maintained judgement accuracy levels over the delay. For fruits_13, again a main effect of delay was found, $F(1,47) = 128.865$, $p < .001$, $\eta_p^2 = .733$, as previously. A main effect of measure was also found, $F(1,47) = 11.188$, $p = .002$, $\eta_p^2 = .192$, again as previously observed for French words_7 and colours_10. No interaction was present, $F(1,47) = 1.333$, $p = .254$, $\eta_p^2 = .028$.

For creatures_17, a slightly different pattern emerges. As previously, a main effect of delay is present in the expected direction, $F(1,47) = 38.788$, $p < .001$, $\eta_p^2 = .452$.

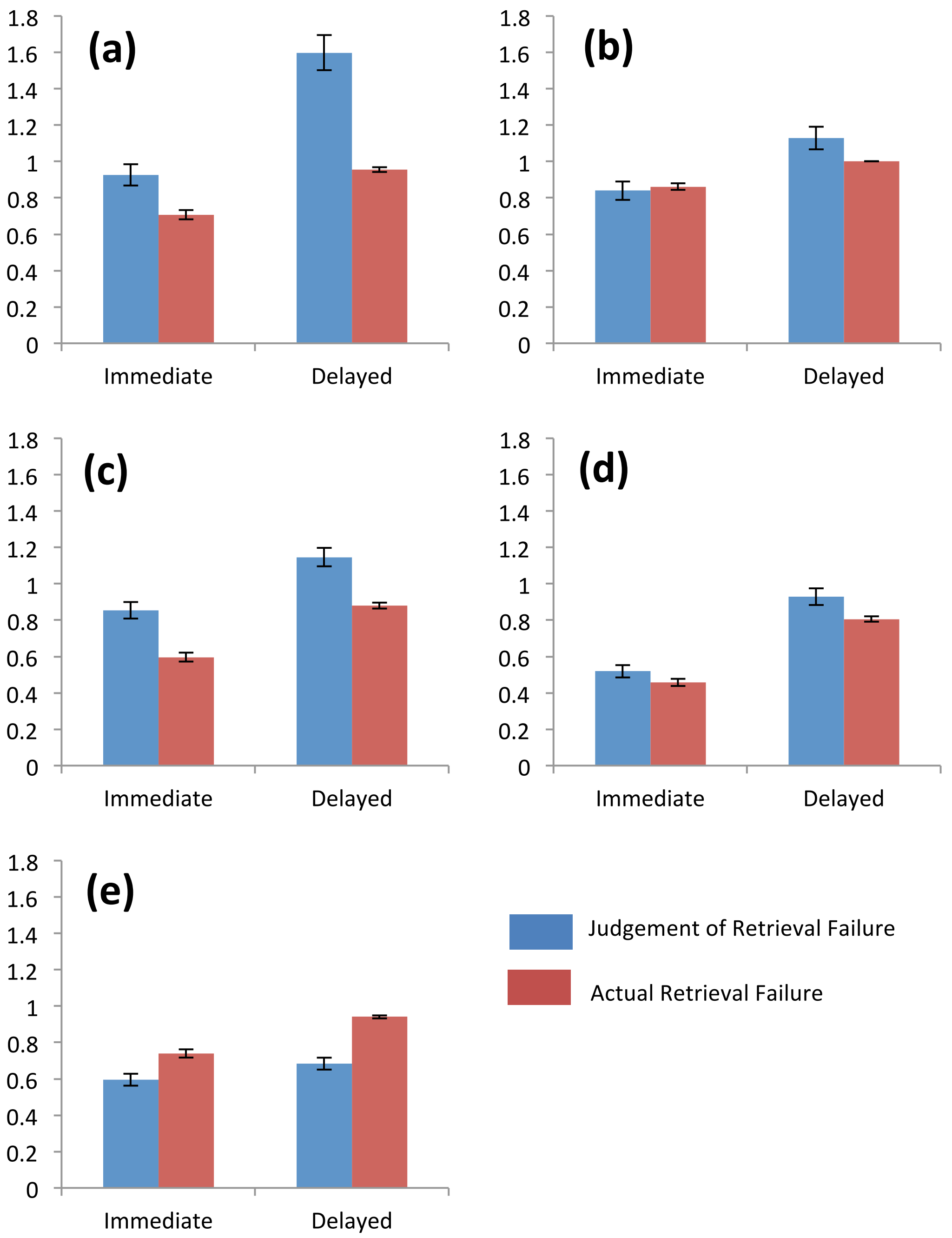


Figure 5.1: Proportion scores for Judgement of Retrieval Failure and Actual Retrieval Failure for (a) French words_7, (b) Nonwords_9, (c) Colours_10, (d) Fruits_13, and (e) Creatures_17 at each delay. Ordinate represents proportion scores. Bars represent the standard error of the mean.

However, the main effect of measure, $F(1,47) = 63.130$, $p < .001$, $\eta_p^2 = .573$, shows that JORF ratings are lower than actual retrieval failure, in contrast to the previous findings. This could be due to the larger set size of creatures. A significant interaction between delay and measure was also present, $F(1,47) = 8.953$, $p = .004$, $\eta_p^2 = .160$. Follow up paired samples t-tests revealed that JORFs remained significantly lower than retrieval failure both immediately, $t(47) = 5.025$, $p < .001$, and after the delay, $t(47) = 7.551$, $p < .001$. However, while retrieval failure increased over the delay, $t(47) = 10.456$, $p < .001$, JORF levels remained stable, with no significant increase or decrease, $t(47) = 2.346$, $p = .022$.

In sum, all categories show a main effect of delay, with JORFs and actual retrieval failure increasing over the 14 week interval (see Table 5.4). All categories except nonwords show a main effect of measure. Figure 5.1 shows that for most materials either immediately or after the delay, the level of JORF is higher than actual retrieval failure (7 out of 10 comparisons). That is, people report that they have failed to retrieve more than they actually have; they overestimate their retrieval failure, or underestimate their memory proficiency. Fruits, colours and French words all show an overestimation of the amount of information that was not recalled, with participants predicting there were more items on each list than were actually nonrecalled. The creatures category showed the opposite effect, with JORFs *lower* than actual retrieval failure. This could potentially be due to the larger category size used rather than being due to the category items themselves. It may be that participants are only able to judge retrieval failure up to a certain size limit. This remains speculative and would require further study to confirm. Interactions between delay and measure were observed for creatures and French words, with accuracy decreasing over the delay, as would be expected.

Table 5.4: Summary of main effects and interactions of 2 (delay) x 2 (measure) ANOVAs for JORF within each category

	French words 7	Nonwords 9	Colours 10	Fruit 13	Creatures 17
Delay	✓	✓	✓	✓	✓
Measure	✓	X	✓	✓	✓
Interaction	✓	X	X	X	✓

Note: ✓ indicates significant effect, X indicates no significant effect, ≈ indicates marginal effect

A Pearson's correlation was run for each category at each delay to further explore the relationship between JORFs and actual retrieval failure (for scatterplots of the data see Appendix C). When tested immediately, all categories except nonwords_9 show a positive correlation between JORF and actual retrieval failure (see Table 5.5). As the number of nonrecalled items increase so do participants' judgements of how many extra items there were on the list that they did not recall, thus showing that the JORF reflects retrieval failure. After a delay, however, no significant correlations were observed. After this time interval it may be that insufficient episodic information is available to participants on which to base the judgement. Without information which is diagnostic of retrieval failure no relationship is observed between the two measures. It should be noted that for nonwords_9 at the delayed test no items were correctly recalled, thus no correlation could be calculated.

Table 5.5: Correlations between Judgements of Retrieval Failure and Actual retrieval failure at each delay

	Immediate		Delayed	
French words_7	0.420	***p* = .003**	-0.074	*p* = .619
Nonwords_9	0.290	*p* = .046	-	-
Colours_10	0.563	***p* < .001**	-0.084	*p* = .569
Fruits_13	0.671	***p* < .001**	0.052	*p* = .726
Creatures_17	0.512	***p* < .001**	0.103	*p* = .484

5.2.2.3 GFOK

GFOKs were elicited by asking participants to predict how many of the items they had failed to retrieve they thought they would later recognise. Thus the GFOK is a proportion of the JORF, limiting the impact of inaccurate JORFs on GFOK resolution. Predictions as to the proportion of unrecalled items that participants would later recognise were compared to the actual proportion of unrecalled items that were recognised in a 2 (delay) x 2 (measure: GFOK and recognition of unrecalled) repeated measures ANOVA (see Figure 5.2). For French words_7, a main effect of delay was found, $F(1,47) = 52.243$, $p < .001$, $\eta_p^2 = .526$. Both measures were higher before the delay than after the delay. A main effect of measure was also found, $F(1,47) = 54.807$, $p < .001$, $\eta_p^2 = .538$. GFOK ratings were lower than actual recognition of nonrecalled items; participants underestimated their ability to correctly recognise items they had failed to retrieve at recall. No interaction between delay and measure was present, $F<1$. For nonwords_9, the same pattern of results was found, with scores higher before the delay than after, $F(1,47) = 97.023$, $p < .001$, $\eta_p^2 = .674$, and participants underestimating their actual recognition, $F(1,47) = 7.455$, $p < .001$, $\eta_p^2 = .137$. Again, no interaction was present, $F(1,47) = 2.046$, $p = .159$, $\eta_p^2 = .042$.

Colours_10 show a slightly different pattern. Here, only a marginal effect of delay was found, $F(1,47) = 6.256$, $p = .016$, $\eta_p^2 = .117$. In addition, no effect of measure was present, $F<1$. Participants were therefore accurate in their GFOK predictions, correctly judging the proportion on non-retrieved items that they would be able to recognise later. No interaction was present, F<1.

For fruits_13, results show the same pattern as those of French words_7 and nonwords_9, with a main effect of delay, $F(1,47) = 51.097$, $p < .001$, $\eta_p^2 = .521$, and a main effect of measure, $F(1,47) = 19.201$, $p < .001$, $\eta_p^2 = .290$, as previously. Also, no interaction between delay and measure was present, $F(1,47) = 2.120$, $p = .152$, $\eta_p^2 = .043$. In contrast, creatures_17 follow a similar pattern to that of colours_10, although the main effect of delay does reach significance for this category, $F(1,47)$

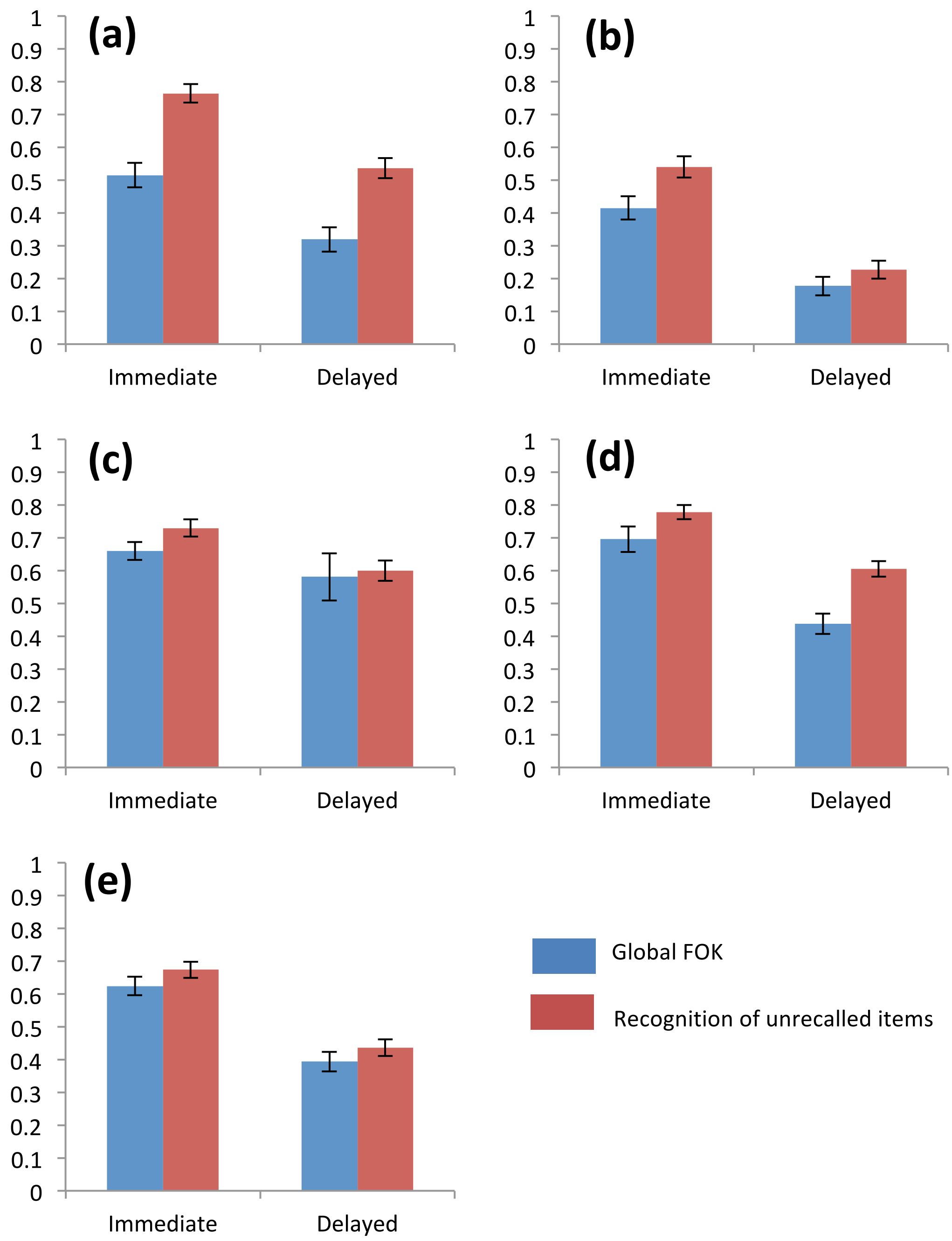


Figure 5.2: Proportion scores for Global FOK and Recognition of Unrecalled Items for (a) French words_7, (b) Nonwords_9, (c) Colours_10, (d) Fruits_13, and (e) Creaetures_17 at each delay. Ordinate represents proportion scores. Bars represent the standard error of the mean.

= 85.983, p < .001, η_p^2 = .647. No effect of measure, $F(1,47)$ = 2.953, p = .092, η_p^2 = .059, or interaction between delay and measure, F<1, was found.

In sum, an effect of delay was observed in all categories, with GFOK and recognition of unrecalled items being greater before the 14 week delay than after the delay, although the effect for colours_10 was marginal (see Table 5.6). As can be seen in Figure 5.2, all 10 comparisons show that actual recognition is higher than predictions of recognition. Similar to the JORF findings, participants appear to underestimate their memory proficiency, judging that they will recognise fewer items than they are actually able to. This underestimation of the GFOK reached significance in three of the five categories; French words_7, nonwords_9 and fruits_13. No interactions between delay and measure were observed, indicating that the delay did not lead to greater impairments in predictive accuracy.

Table 5.6: Summary of main effects and interactions for 2 (delay) x 2 (measure) ANOVAs for GFOK within each category

	French words 7	**Nonwords 9**	**Colours 10**	**Fruits 13**	**Creatures 17**
Delay	✓	✓	≈	✓	✓
Measure	✓	✓	X	✓	X
Interaction	X	X	X	X	X

Note: ✓ indicates significant effect, X indicates no significant effect, ≈ indicates marginal effect

The relationship between GFOK and recognition of unrecalled items was analysed using Pearson's correlations for each category at each delay (see Table 5.7). However, no relationships were observed except for within the creatures_17 category and only following the delay. This indicates that judgements of future recognition are not associated with actual recognition performance.

Table 5.7: Pearson's correlations for Global FOK judgements and recognition of unrecalled items for each category at each delay

	Immediate		Delayed	
French words_7	0.254	*p = .082*	0.224	*p = .127*
Nonwords_9	0.206	*p = .159*	-0.029	*p = .847*
Colours_10	0.093	*p = .531*	-0.119	*P = .422*
Fruits_13	0.045	*p = .762*	0.107	*p = .467*
Creatures_17	-0.105	*p = .479*	0.371	***p = .009***

5.2.3 Discussion

The present study aimed to establish the feasibility of two new metacognitive measures, the judgement of retrieval failure and the global FOK. In addition, it also aimed to assess their sensitivity to a classic memory manipulation of retention interval. Effects of delay on recall and recognition were as expected, with both decreasing over time.

The effect of delay on memory performance is detected with the JORF measure. All categories showed an increase in JORF and actual retrieval failure, thus as participants failed to recall more items, their estimate of the remaining items also increased. The pattern of over or under estimation did show some variation depending on category. Fruits, colours and French words all showed an overestimation of the number of items which had not been recalled. Participants were therefore underconfident in their memory ability, as they believed that they had recalled a smaller proportion of the items on the list than was actually the case. In contrast, for creatures participants were overconfident, and judged there to be fewer items remaining than there actually were. It was only for nonwords that no difference was observed between JORF and retrieval failure, indicating that participants were able to accurately judge the amount of information they had failed to retrieve.

It is interesting to note the lack of consistency with regards to under or overconfidence in the JORF predictions obtained for each category. Only three categories show the same pattern, and even this may be considered at odds with what would be expected from research looking at estimates of accurate recall (Bunnell, Baken, & Richards-Ward, 1999; Hertzog, Saylor, Fleece, & Dixon, 1994). Here, when participants are asked to judge the number of words they have correctly recalled, judgements are either accurate or slightly overestimate their true performance. It may therefore be expected that judgements of retrieval failure would mirror these effects, with a slight underestimation of the number of items not recalled from the list. This is indeed what is shown in the creatures category, where estimates are lower than actual retrieval failure.

There are three possible explanations for the different pattern observed within the fruits, colours and French words. First, the judgements of recall obtained by Bunnell et al. (1999) and Hertzog et al. (1994) informed participants of the total number of items that had been presented. Perhaps by giving this detail it affected the referent used to judge performance against. With the JORF no such referent was given, thus participants had only their assessment of memory to base their predictions on. Second, the framing of the question asked between judgement of recall and the JORF paradigm are different. Evidence from the decision making literature has shown that the framing of a question can have a great impact on people's judgements, in some cases even reversing their decisions (Tversky & Kahneman, 1981). There is emerging evidence within the metacognitive literature that the framing of JOLs also affects the judgements made. Finn (2008) observed that judgements of the chance of forgetting the second word of a pair when given the first word were much closer to actual recall performance than judgements of the chance of remembering the second word of a pair when given the first word (but see Serra & England, 2012, for an alternative view). Thus by asking participants to judge how many items there were remaining on the list rather than how many they had recalled from the list this may have affected their assessment.

The third explanation of why overestimation occurs is linked to the accuracy of judgements found in the nonwords category and the mechanisms behind the reporting of false memories. Nonwords are completely novel items, and do not belong to a larger semantic category. Thus unlike fruits, colours, creatures and even French words to some extent, it is not possible to recall exemplars and then assess whether they were present or not (Dodson & Schacter, 2001, 2002). This may be the source of the inaccuracy within the fruits, colours and French words categories. By providing the category name this may automatically activate exemplars from the category rather than only those items which were presented at encoding (Gallo et al., 2001). This may therefore increase the representation of the category size the participant has, leading to interference and an overestimation of the number of items that have not been recalled. Alternatively, according to fuzzy trace theory, the gist memory trace may be activated while the verbatim trace is not, leading to non-presented exemplars being identified (Brainerd & Reyna, 2002; Gernsbacher, 1985). For nonwords, no such exemplars can be activated, thus the representation remains independent of any interference effects. This explanation is however problematic due to the observed underestimation in the creatures category. It may be possible that the larger sample size used for this category may be interacting with the interference effects. Participants may only be able to make JORF estimates up to a certain category size, after which the representation becomes too imprecise to accurately judge. This provides an interesting avenue for further research, to potentially establish at what point individuals are unable to access diagnostic information and give up trying to make estimations of their memory ability.

The presence of interactions with delay for some categories and the pattern of correlations and delay is also of interest. It could be expected that delay would interact with JORF and actual retrieval failure due to the decay of memory over time. As the delay increases, the memory trace decreases, thus the representation of the category would presumably also decrease leading to less information on which to base a judgement (Gernsbacher, 1985). This would therefore lead to a decrease in accuracy after a delay. Where interactions are present within the data, this is the pattern of results observed (creatures and French words). Likewise, the correlations suggest that immediate JORFs are closely associated with memory performance, whereas JORFs following the delay are not. This effect can be explained by fuzzy trace theory (Brainerd & Reyna, 2002). With the immediate JORF, the verbatim memory trace of each item provides the basis of the metacognitive judgement, whereas following the retention interval this verbatim trace will have decayed, leading to the JORF being primarily based on the gist memory trace which all items share. Thus the verbatim traces provide partial information relevant to each individual item giving greater diagnostic detail, whereas the gist trace provides partial information shared by each item and so is not as diagnostically useful when assessing retrieval failure.

The GFOK measure was also sensitive to the effects of delay, with both predictions of future recognition and actual recognition decreasing following the retention interval. This follows the same pattern as item-by-item FOKs as examined in Experiment 3.2, where retention interval led to a decrease in the magnitude of the FOK ratings. There is also evidence of underconfidence in the GFOK, with predictions being significantly lower than actual recognition performance for fruits, creatures, French words and nonwords. Interestingly, no interactions with delay were observed, again in agreement with the item-by-item FOKs in Experiment 3.2. Thus GFOKs appear to be affected by retention interval in a similar way to item-by-item FOKs, adding weight to the assumption that the GFOK and standard FOK are measuring the same concept.

The findings of Experiment 5.1 provide validation of the JORF as a novel measure of global metacognitive awareness. It has been shown to be sensitive to the memory manipulations of retention interval and list size. The type of memory can also affect the magnitude and accuracy of metacognitive judgements, as already discussed in the context of FOKs and ageing in Chapter 2. The observation that JORFs for nonwords were closer to actual retrieval failure than for those of semantic categories could indicate a similar effect of memory type for the JORF measure. Experiment 5.2 therefore aimed to assess the accuracy of JORFs for semantic and episodic memory separately. To achieve this, the semantic test will be to recall elements from the periodic table and judge the number of elements which have not been recalled. Participants were again Psychology undergraduates, thus were required to have some background in science to gain entry on the course and therefore will have some knowledge of the periodic table. The episodic task was a selection of items from Experiment 5.1. Although these items belonged to separate semantic categories, in this instance they were presented as a single list and participants were not prompted at recall to report the items according to semantic category. Therefore the influence of semantic memory should be minimised.

5.3 Experiment 5.2: Judgements of retrieval failure for semantic and episodic memory

5.3.1 Method

5.3.1.1 Participants

A total of 52 Psychology undergraduate students from the University of Leeds (*M* age = 21.1 years, *SD* = 1.38) were recruited for the study.

5.3.1.2 Materials and procedure

Episodic Task

Participants were presented with 32 words from three categories: fruits (13 items), colours (10 items) and nonwords (nine items). Items were presented simultaneously on a single Powerpoint slide using Helvetica font size 25. The slide was projected onto a lecture theatre screen for two minutes. Following presentation, a response sheet was provided to give recall and JORF predictions. Participants were asked to write down as many of the words presented as possible. Two judgements were then obtained. First, participants were asked to state how many other items were on the list that they had not recalled. This provided a judgement of retrieval failure for the entire list irrespective of category type. Following this, participants were asked to judge how many other nonwords only were on the list that they had failed to recall. This provided a more fine-grained estimate of retrieval failure but still avoiding the one-to-one cueing inherent in the FOK paradigm.

Semantic Task

Upon completion of the episodic task, a semantic version of the task was administered. No learning trial was given. Participants were provided with a response sheet and asked to recall as many elements of the periodic table as they could. This gave a potential set size of 118 items. As previously, upon completion of recall participants were asked to judge how many other elements there were in the periodic table that they had not been able to retrieve. Finally, participants were also asked to judge how many elements there were beginning with the letter Z in addition to any they had recalled.

5.3.2 Results

5.3.2.1 Memory

Due to differences in category size, proportions were calculated for all measures, as for Experiment 5.1. Recall was calculated by dividing the number of correct responses by the total number of possible responses i.e. category set size. Recall accuracy for the Semantic and Episodic tasks were compared using a paired samples t-test. Participants recalled a significantly higher proportion of the episodic materials than the semantic materials, $t(51) = 9.964$, $p < .001$.

To establish whether participants were more or less able to judge their retrieval failure for semantic or episodic items, the proportionalised data were submitted to a 2 (memory type: Semantic and Episodic) x 2 (measure: JORF and actual retrieval failure) repeated measures ANOVA (see Figure 5.3). No effect of memory type was found, $F(1,51) = 1.971$, $p = .166$, $\eta_p^2 = .037$. A main effect of measure was present, $F(1,51) = 58.945$, $p < .001$, $\eta_p^2 = .536$. Judgements of retrieval failure were lower than actual retrieval failure, indicating participants underestimated the number of items that they have failed to recall. A significant interaction was also present,

$F(1,51) = 43.946$, $p < .001$, $\eta_p^2 = .463$. T-tests revealed that for semantic materials, JORFs were significantly lower than actual retrieval failure, $t(51) = 9.340$, $p < .001$. However, for the episodic materials, no difference was found between JORF and actual retrieval failure, $t(51) = 0.555$, $p = .581$. This may indicate that participants were able to accurately monitor the proportion of unrecalled items for the episodic materials but not for the semantic materials.

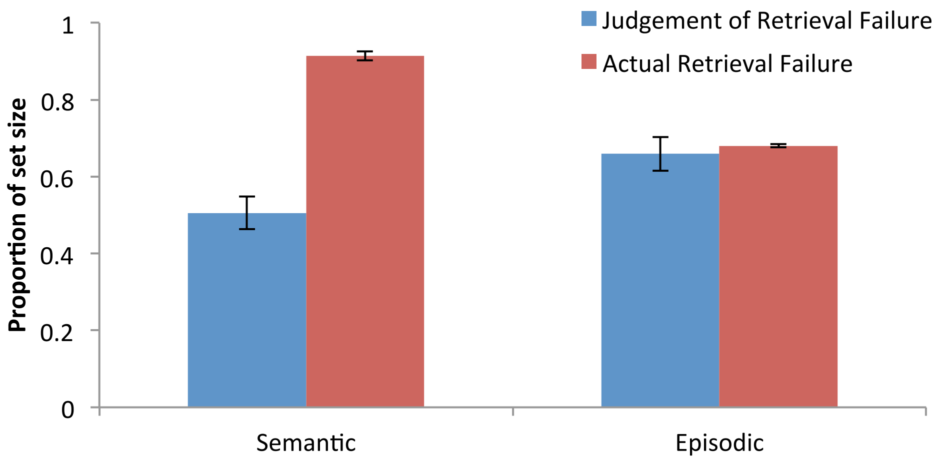


Figure 5.3: Mean JORF and actual retrieval failure for the semantic and episodic tasks.

Pearson's correlations were also used to examine the relationship between JORF and actual retrieval failure for semantic and episodic memory (see Figure 5.4). No relationship was found for the semantic items, $r(52) = -.203$, $p = .149$. As a group, participants were unable to adjust their judgement according to actual remaining set size. In contrast, a moderate positive correlation was found between JORF and retrieval failure for episodic items, $r(52) = .535$, $p < .001$. As the proportion of nonrecalled items increases, so do participants' judgements of the number of items they have failed to retrieve from the list.

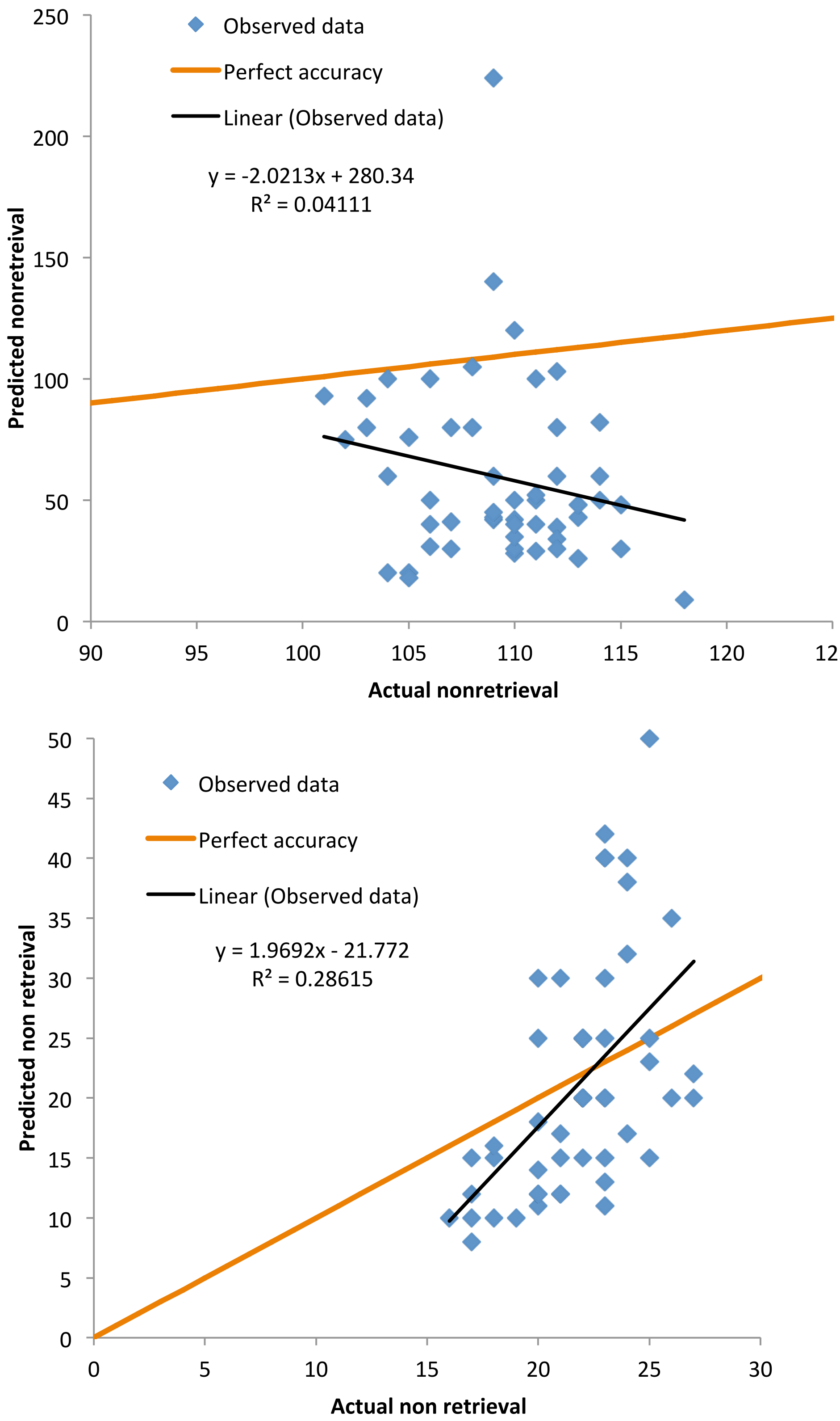


Figure 5.4: Scatterplot of predicted non retrieval and actual nonretrieval for (a) semantic and (b) episodic items.

To further examine the differences between semantic and episodic judgements, a subset of each was also analysed. For this purpose the accuracy and predictions for elements beginning with the letter 'z' were chosen from the semantic items, and the accuracy and judgements for nonwords only was chosen from the episodic items. Again, the proportionalised data were analysed using a 2 (memory type) x 2 (measure) repeated measures ANOVA (Figure 5.5). No effect of memory type was found, $F(1,51) = 1.420$, $p = .239$, $\eta_p^2 = .027$. A main effect of measure was again observed, $F(1,51) = 27.283$, $p < .001$, $\eta_p^2 = .349$. In contrast to the whole data set, the present analysis found that JORFs were greater than actual retrieval failure. Thus for this smaller subset of items participants significantly overestimated the remaining items they had failed to recall. An interaction between memory type and measure was also present, $F(1,51) = 37.937$, $p < .001$, $\eta_p^2 = .427$. As previously, for the semantic items a significant difference was found between JORF and actual retrieval failure, $t(51) = 5.980$, $p < .001$, with participants judging there to be more items that they had not recalled than were actually remaining. However, participants were able to accurately assess how many episodic items they had failed to retrieve, with no difference found between JORF and retrieval failure, $t(51) = 1.393$, $p = .170$.

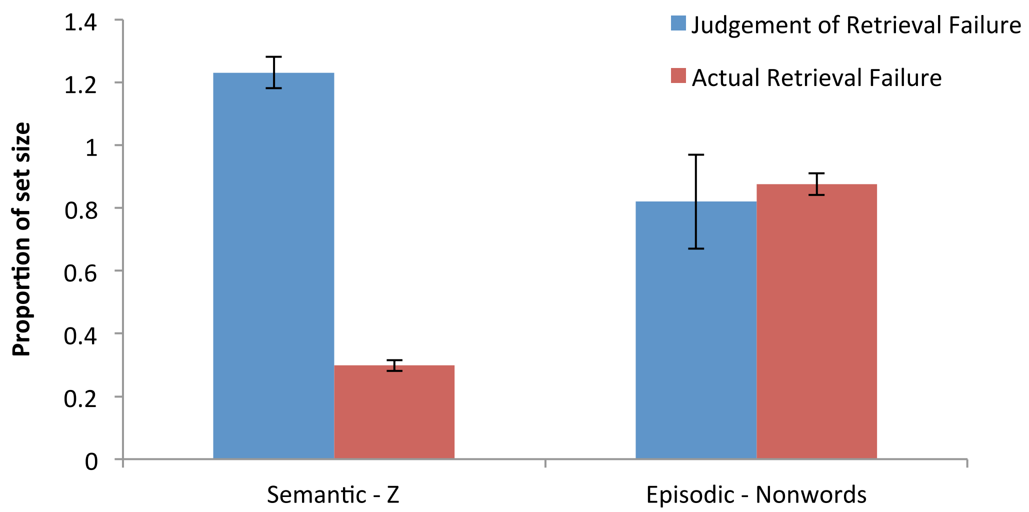


Figure 5.5: Mean JORF and retrieval failure for subset of items from the episodic and semantic tasks

Again, Pearson's correlations were examined to establish if any relationship could be observed between JORF and actual retrieval failure. In this instance, no correlation was found within semantic items, $r(52) = -.060$, $p = .670$, nor episodic items, $r(52) = .111$, $p = .434$. This could be due to the small set sizes used, which will not only increase the amount of error but also lead to a restricted range of responses being examined. This is of particular concern for the semantic subset.

5.3.3 Discussion

The aim of the current study was to compare the JORF for semantic and episodic memory. As for Experiment 5.1, the key analysis was the presence or absence of an interaction between memory type and measure. Results showed episodic JORFs to be accurate, with no difference between judgements of retrieval failure and actual retrieval failure. In contrast, a significant difference was found for semantic JORFs, with judgements much lower than actual retrieval failure. That is, participants underestimated how many elements there were. In addition, a clear correlation between retrieval failure and JORF was observed for episodic items, whereas no relationship was found for semantic items. This could imply that JORFs are affected by the type of memory tested i.e. semantic or episodic, as other metacognitive judgements such as the FOK are affected. Alternatively, this could be an artefact of the design used, with different set sizes affecting the ability to monitor retrieval failure. For the episodic task, the information is provided on which to base the JORF: although participants are not asked to count the number of items presented, the total set is given and may be used to inform the judgement. For the semantic task, each participant may have a different concept of what the total set size is. With such a large set size, this concept has greater possible variability than a smaller set size would. Thus the greater variability will lead to greater error, reducing the accuracy of the JORF. A replication study with materials of a similar set size would provide a clearer interpretation of the present data.

It is interesting to note that, unlike Experiment 5.1, the episodic JORF obtained here shows no effect of measure; JORFs were similar to actual retrieval failure. This could be due to the lack of semantic categorisation given in the episodic task within the present study. Although the individual items belonged to the categories of fruits, colours and nonwords, they were presented as a single list, thus participants were not prompted to encode or retrieve them within a particular semantic structure. Participants may not therefore have been aware of the semantic groupings, which may have influenced their strategy at recall and JORF. The generate-recognise strategy (Anderson & Bower, 1972) relies on some kind of grouping cue to allow generation of appropriate exemplars. Errors occur due to falsely recognising non-presented exemplars which are semantically related to items presented at encoding (Johnson et al., 1993). In Experiment 5.1, this strategy is encouraged by giving category names as the cue for memory. This could lead to a higher incidence of retrieval errors being reflected in errors at JORF, driving the main effect of measure observed.

In contrast, within Experiment 5.2 no explicit mention of category grouping was made; participants were instructed to recall and judge 'the list'. Thus participants may not have spontaneously grouped the items presented by category, and may not have subsequently used the category groupings in order to adopt a generate-recognise strategy. This would therefore reduce the risk of misidentifying semantically related items as being present at encoding, leading to lower JORFs which are more similar to actual retrieval failure. This explanation would be easily verifiable in a replication study by asking participants whether they were aware of the category groupings and also to report on their strategy use following completion of the experiment.

The results obtained for the semantic category could either be considered as reflecting a true deficit in resolution for semantic materials compared to episodic materials, or could be due to constraints imposed by the task used. First, it is worth considering which category is used to obtain the assessment of semantic JORF. It is important to choose a category where some level of vagueness regarding the total number of items exists. If the complete set size is known independently by the majority of people this will enable participants to give perfectly accurate JORFs. For example, the majority of people will know the number of dwarves in the fairy tale of Snow White, even if they cannot recall each dwarf by name. No judgement would actually be necessary in this case; the participant can simply subtract recall from the known number of items regardless of subjective experiences related to failed retrieval of individual items. The task would then be measuring knowledge rather than metacognitive evaluations. The knowledge of the total set size would also drive memory behaviour, as the participant will have a concrete target number of items to retrieve. However, there must also be a stable finite number of items that constitute the category to allow for analysis of the representativeness of the JORF. Categories such as countries within the EU are therefore not suitable as the total members of the semantic group alter regularly. Elements of the periodic table provide a balance between these two extremes, with a fixed number of elements composing the category but with a suitable level of vagueness to prevent participants reporting knowledge rather than giving judgements.

If the JORF is not based on knowledge of the total set size, what then is it based on? The generate-recognise strategy cannot be employed with semantic items; participants are required to generate only in order to complete the task. Thus they must first identify all the known exemplars from the category, which are then reported at the recall test. The estimate of the remaining items is then based on the number of "known unknowns" (Glucksberg & McCloskey, 1981; Hampton, Aina, Andersson, Mirza, & Parmar, 2012), items which are not retrievable from memory but that the rememberer believes must exist. The JORF would therefore represent the limits of the awareness of ignorance, how much is known about what is not known. Error between the JORF and actual retrieval failure arises due to unknown unknowns (Hampton et al., 2012), items which have not been recalled and that the rememberer has no awareness of existing.

The unknown unknowns could therefore lead to the observed differences between JORF and retrieval failure in the semantic task used here. These unknown unknowns would not have an influence within the episodic task, as the participant is presented with all relevant information to encode. The number of unknown unknowns subsequently leads to greater discrepancy between the semantic JORF and semantic retrieval failure than is observed between the episodic JORF and the episodic retrieval failure. This would suggest a deficit in semantic JORF ability. However, as already noted, the semantic JORF is highly reliant on the task used. If the total set size of the semantic category is known independently of category exemplars, then semantic JORFs would show greater similarities between the two measures than episodic JORFs.

This lack of awareness of ignorance relates to participant expertise. Psychology undergraduate students were used in the study who will have had exposure to the periodic table through their secondary school teaching. However they are not experts within the field of chemistry, and therefore will have limited semantic knowledge of the subject. If the study were to be repeated with chemistry undergraduate students, the results may be very different. This would also have implications more generally for the role of expertise in metacognitive judgements. As discussed in Chapter 2, expertise is not thought to influence FOK resolution. However, it would seem logical that expertise would affect JORF resolution. This would provide an interesting avenue of further investigation to establish if and why expertise may play a role in some metacognitive judgements but not others.

5.4 Experiment 5.3: Judgements of retrieval failure and global FOKs in early stage dementia

5.4.1 Introduction

A lack of awareness of cognitive impairments is one of the clinical features of Alzheimer's disease (AD), emerging in the early stages and becoming more pronounced as the disease progresses (Feher, Mahurin, Inbody, Crook, & Pirozzolo, 1991). Although the domain in which this lack of awareness, or anosognosia, occurs can vary, the most common area of deficit is within the memory domain (Barrett, Eslinger, Ballentine, & Heilman, 2005). As well as being of diagnostic importance, anosognosia can have profound impacts on prognosis, as it may lead to delays in seeking treatment and also affect compliance with treatment options (Clare & Woods, 2004b). If the patient is not fully aware of where their memory failings lie, they will not implement strategies to aid recall.

There has been extensive examination of the impact of AD on memory awareness in the metacognitive literature (for reviews see Pannu & Kaszniak, 2005; Souchay, 2007). For global judgements, AD patients typically overestimate their memory ability. In one of the first studies to examine metacognition in AD, Schacter, McLachlan, Moscovitch, & Tulving (1986) examined predictions of recall performance in healthy individuals and in patients with memory impairments due to closed head injury, ruptured aneurysm, or AD. Compared to the healthy control group, all patient groups recalled fewer items. However, while the closed head injury and ruptured aneurysm patient groups accurately predicted the amount of items they would recall, the AD group judged that they would remember more items than they actually did. Although all patient groups exhibited a memory impairment, this was not in itself sufficient to impair memory predictions; some unique deficit occurs within the AD patients when judging future performance. Barrett et al. (2005) observed a deficit in retrospective judgements of memory ability within AD patients, again with patients overestimating their performance on the task. Ansell & Bucks (2006) asked participants to predict performance both pre- and post-testing. Again, when compared to healthy controls patients with AD overestimated their memory ability both pre- and post- test, although accuracy was greater following the recall test than prior to.

Although the previous studies show a deficit in the accuracy of global judgements in AD, patients have also been shown to be sensitive to certain aspects of memory tasks. Indeed, the findings of Ansell & Bucks (2006) described above suggest that AD patients are able to use their experience of completing the memory task to adjust their judgements down to closer reflect their performance, albeit not to perfect accuracy. Moulin et al. (2000a) similarly observed greater accuracy in AD patients with task experience, a finding which was also replicated by Schmitter-Edgecombe & Seelye (2011). In addition, AD patients have been shown to be sensitive to the effects of delay (McGlynn & Kaszniak, 1991), repeated learning (Duke, Seltzer, Seltzer, & Vasterling, 2002; though see also Moulin, Perfect, & Jones, 2000b), type of memory test (Moulin, 2002) and item characteristics (Moulin, Perfect, & Jones, 2000c). The sensitivity approach (Moulin, 2002) demonstrates that, although patients are not able to monitor the online aspects of memory, thus leading to overconfidence, they still have access to general knowledge (or theory based aspects, Koriat et al., (2004)) of memory function and can adjust their predictions according to these known rules.

Based on the findings from previous examinations of global judgements in AD, it may be expected that patients will not show a high degree of accuracy when making JORFs, though they will show sensitivity to the level of failed retrieval. Thus categories where they have recalled fewer items will receive higher JORFs than categories where they have recalled a larger number of items. Due to the pattern of results in Experiment 5.1, it could be considered that the GFOK resembles item-by-item FOKs. Thus examination of the FOK literature in AD may give an indication of the ability of AD patients to make this judgement. The primary focus of FOKs in AD has been in semantic memory, establishing monitoring ability for general knowledge. Here, despite deficits in recall ability, no impairment of FOK accuracy has been found (e.g. Bäckman & Lipinska, 1993; Lipinska & Backman, 1996). Episodic FOKs have received little examination, either due to the high volume of items required to enable measurement of the FOK, leading to floor performance in AD patients (Souchay, 2007) or due to overly complex ratings scales which patients failed to use appropriately (Pappas et al., 1992). Souchay, Isingrini, & Gil (2002) used a simpler Yes/No FOK task to assess episodic monitoring ability with a reduced number of items compared to typical FOK tasks, thus enabling FOK assessment. It was found that AD patients FOK accuracy was at chance performance; their ratings were not predictive of future recognition. Indeed, AD patients made a similar proportion of Yes FOK judgements, despite reduced recognition accuracy, indicating a level of overconfidence in their memory ability. It may therefore be expected within the present study that the GFOK measure may also show a lack of predictive accuracy within the AD patient group, with participants predicting that they will recognise a larger proportion of nonrecalled items than they actually do.

5.4.2 Method

5.4.2.1 Participants

Two groups of participants were recruited for the study: 16 patients diagnosed with dementia (AD) and 16 older adult controls (OAC). Diagnosis of dementia was made by a clinician using neuropsychological examination, MMSE, family interview and medical examination. Patients with a history of stroke were excluded from the study. OACs were recruited from the local community and were screened with the MMSE prior to the study. Participant characteristics can be seen in Table 5.8. No differences were observed between groups for age, $t(30) = 1.482$, $p = .149$, or for years of formal education, $t(30) = 1.147$, $p = .260$. OACs did score higher than the AD group on the National Adult Reading Test (NART), $t(30) = 2.057$, $p = .049$.

Table 5.8: Means (*SD*) of age, MMSE, education and NART scores for patients diagnosed with dementia and older adult control participant groups

	Dementia patients	Older adult controls
Age	83.06 (4.31)	80.81 (4.28)
MMSE	24.31 (2.57)	28.50 (0.89)
Years of formal education	12.00 (1.32)	12.81 (2.51)
NART	107.40 (31.59)	124.87 (12.20)

5.4.2.2 Materials

Materials were based on those used in Experiment 5.1 and 5.2 with a reduced number of items per set. Participants were randomly assigned to one of two versions of the stimulus set. In version A, participants were presented with eight fruits, followed by four French words, five colours, eleven creatures, and three nonwords. In version B, participants were presented with five creatures, eleven fruits, three French words, four nonwords and eight colours. Piloting with three AD patients indicated that if more than four French words or nonwords were given participants were not motivated to attempt to learn the items, perceiving the task as too difficult.

5.4.2.3 Procedure

Participants were tested individually, and were informed that they would be asked to learn five lists of words for a later recall test. Lists were presented on Powerpoint slides using a Dell laptop with 17" screen. Each Powerpoint slide comprised of the category exemplars given in Calibri font size 52. For each list, participants were first asked to name the category to which the words belonged[7]. Following their response, the category name was presented at the top centre of the slide in bold and the timed encoding for the list would begin. Each list was presented for four seconds per item i.e. lists of length 3 items would be presented for 12 seconds. During presentation, participants were asked to read each word aloud to ensure all items were attended to. Once all items had been read aloud participants were asked to spend the remaining time the words were on the screen to try and learn them.

Following presentation of all five lists, the recall test began. Participants were given the category name and asked to recall as many exemplars as they could from the list they had been asked to learn without guessing. Responses were given verbally, and the experimenter recorded each response. Once participants reported all they could recall, participants were asked to state how many other items were on the list (the JORF), and also to predict how many of those other items they would recognise if shown them (global FOK). After each judgement had been made, the next category name was given and recall, JORF and global FOK were obtained for that category. Category prompts were given in the same order as categories were presented at recall. Following completion of the recall stage, an old / new

---

[7] All participants were able to give the correct category name without prompting

recognition task was given. Participants were shown a printed sheet listing exemplars for each category, and were asked to circle or verbally report which items they recognised as being from the study list. The number of distracters matched the number of target items. Again, categories were presented in the same order as during learning, and participants were discouraged from guessing.

5.4.3 Results I: Analysis by category label

As previously, each word list category was analysed separately due to differences in word characteristics. To allow items to be collapsed across version, measures of the number of items correctly recalled, discrimination index and JORF were divided by list length giving a proportion of category size. The global FOK measure was divided by JORF to give a proportion of assumed remaining set that would be recognised, as previously. Alpha was set at 0.01 to control for multiple comparisons

5.4.3.1 Memory

First, it was important to establish whether the AD group showed the expected memory impairments with the present task. Independent samples t-tests of recall accuracy in all categories confirms this, as OACs recalled a greater proportion of items than AD patients for fruits, $t(30) = 4.006$, $p < .001$, $d = 1.43$; colours, $t(30) = 3.602$, $p < .001$, $d = 1.27$; creatures, $t(30) = 6.589$, $p < .001$, $d = 2.44$; French words, $t(30) = 4.82$, $p < .001$, $d = 1.75$; nonwords, $t(30) = 4.666$, $p < .001$, $d = 2.00$ (see Figure 5.7 for means of retrieval failure i.e. 1 – recall). Recognition accuracy was examined with independent t-tests of the discrimination index (hits minus false alarms; see Figure 5.6). Again, OACs showed greater accuracy than AD in all categories (see Figure 5.6 for means); fruits, $t(30) = 5.191$, $p < .001$, $d = 1.90$, colours, $t(30) = 5.645$, $p < .001$, $d = 1.98$, creatures, $t(30) = 4.112$, $p < .001$, $d = 1.51$, nonwords, $t(30) = 6.640$, $p < .001$, $d = 2.60$, French words, $t(30) = 6.661$, $p < .001$, $d = 2.77$.

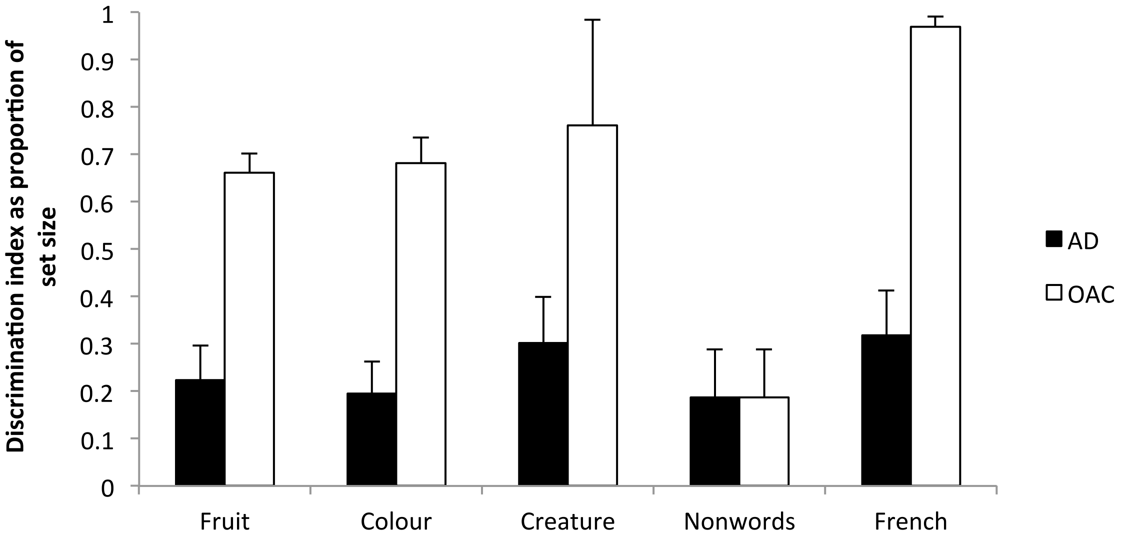


Figure 5.6: Discrimination index for Alzheimer's Disease (AD) patients and older adult controls (OAC) for each category type

The number of intrusion errors at recall was also analysed (see Table 5.9). No differences were observed (fruits, $t(30) = 1.907$, $p = .069$, $d = 0.70$; colours, $t(30) = 1.506$, $p = .149$, $d = 0.59$; creatures, $t(30) = 0.434$, $p = .667$, $d = 0.16$; French words, $t(30) = 1.054$, $p = .300$, $d = 0.39$; nonwords, $t(30) < 0.000$, $p = 1$,).

Table 5.9: Mean (SD) intrusion errors within each category type made by AD and OAC groups

	Fruits	Colours	Creatures	Nonwords	French words
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
AD	1.313 (1.250)	1.875 (2.964)	0.563 (0.964)	0.188 (0.403)	0.063 (0.250)
OAC	0.625 (0.719)	0.688 (1.078)	0.438 (0.629)	0.188 (0.403)	0.188 (0.403)

5.4.3.2 JORF

The JORF was compared to actual retrieval failure in a 2 (group: AD and OAC) x 2 (measure: JORF and retrieval failure) mixed ANOVA for each category (see Figure 5.7). For fruits, a main effect of group was found, $F(1,30) = 15.204$, $p = .001$, $\eta_p^2 = .336$, with OAC lower than AD. No effect of measure was found, $F(1,30) = 1.960$, $p = .172$, $\eta_p^2 = .061$, nor was an interaction present, $F(1,30) = 1.092$, $p = .304$, $\eta_p^2 =$

.035. A similar pattern was present for colours, although the effect of group was marginal, F(1,30) = 5.998, p = .020, η_p^2 = .167, with no effect of measure, F(1,30) = 1.694, p = .203, η_p^2 = .053, and no interaction, F<1. For creatures, a main effect of group was found, F(1,30) = 2.812, p < .001, η_p^2 = .443, with OAC scoring lower than AD. No effect of measure, F<1, or interaction, F<1, was present. French words also follow the same pattern, with a main effect of group, F(1,30) = 20.252, p < .001, η_p^2 = .403 indicating lower overall ratings in OAC than AD, no effect of measure, F<1, and no interaction, F(1,30) = 2.110, p = .157, η_p^2 = .066.

In sum, a clear effect of group was found, with AD patients showing greater retrieval failures and JORFs than OACs, as expected (see Table 5.10). Interestingly, no main effect of measure was observed for any category. This suggests that, overall, JORFs were similar to actual retrieval failure. As can be seen in Figure 5.7, for the OAC group especially the JORF and retrieval failure are at similar levels. The AD group appear to show a level of under confidence in the fruit category, with overconfidence for the nonwords and French words, however these differences are not detectable within the analysis. The key examination of the interactions also failed to reach significance in any category. Therefore, both AD and OAC groups were able to give JORF assessments which were similar to the actual retrieval failure experienced. Although patients with dementia failed to retrieve more items, they also *judged* that they had failed to retrieve more items.

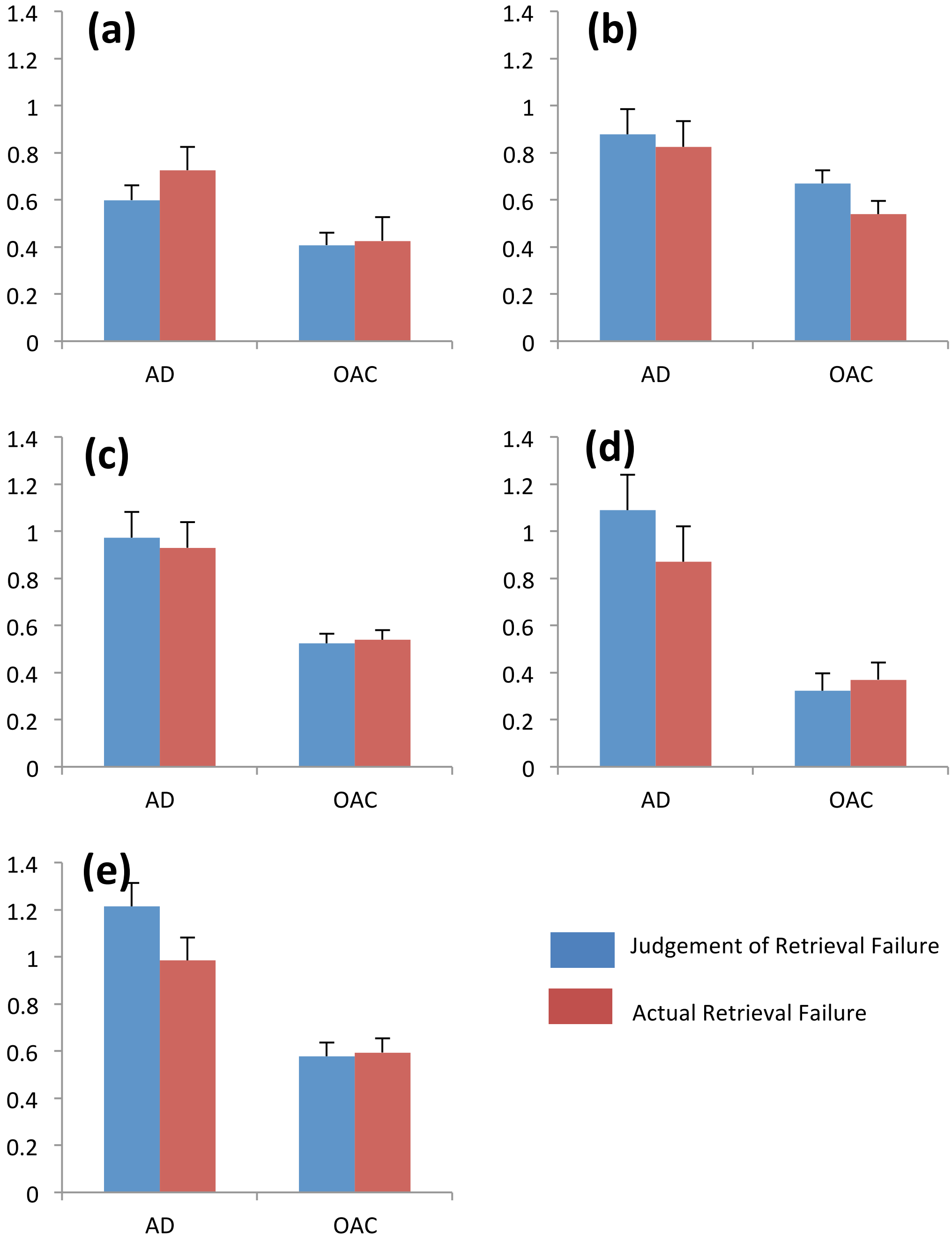


Figure 5.7: Proportion scores for Judgement of Retrieval Failure (JORF) and Actual Retrieval Failure (a) Fruits, (b) Colours, (c) Creatures, (d) French words, and (e) Nonwords in the dementia (AD) group and Older Adult Control (OAC) group. Ordinate represents proportion scores. Bars represent the standard error of the mean.

Table 5.10: Summary of main effects and interactions for 2 (group) x 2 (measure) ANOVAs of JORF within each category

	Fruit	Colours	Creatures	French words	Nonwords
Group	✓	≈	✓	✓	✓
Measure	X	X	X	X	X
Interaction	X	X	X	X	X

Note: ✓ indicates significant effect, X indicates no significant effect, ≈ indicates marginal effect

The relationship between JORF and retrieval failure was further explored with Pearson's correlations between the two measures within each group (Table 5.11). Although no significant difference was observed between measure and JORF, few correlations reach significance. The exceptions are those for OACs within the French words and Nonwords categories, where strong positive correlations are observed. This suggests that, despite the similarity between the JORF and retrieval failure, they do not generally vary with each other (see Appendix D for scatterplots).

Table 5.11: Pearson's correlations between JORF and actual retrieval failure for each participant group

	AD		**OAC**	
	r	***p***	**r**	***p***
Fruits	-.133	*.624*	.582	*.018*
Colours	.430	*.097*	.484	*.057*
Creatures	.103	*.703*	.260	*.332*
French words	.512	*.043*	.848	***<.001****
Nonwords	-.326	*.218*	.884	***<.001****

Note: N = 16 for AD and N = 16 for OAC. * indicates correlation is significant at alpha level of 0.01.

5.4.3.3 Global FOK

Global FOK judgements were compared to the proportion of nonretrieved items which were successfully recognised in a series of 2 (group) x 2 (measure: global FOK and recognition) mixed ANOVAs. For the fruit category, a marginal effect of group was found, $F(1,30) = 7.277$, $p = .011$, $\eta_p^2 = .195$, with OACs giving higher scores than the AD group. No effect of measure, $F(1,30) = 1.782$, $p = .192$, $\eta_p^2 = .056$, or interaction between group and status, $F<1$, was found. For colours, a main effect of group was found, $F(1,30) = 9.502$, $p = .004$, $\eta_p^2 = .241$. No effect of measure, $F(1,30) = 4.024$, $p = .054$, $\eta_p^2 = .118$, , or interaction, $F(1,30) = 3.725$, $p = .063$, $\eta_p^2 = .110$, was present. The creatures category followed the same pattern. A main effect of group showed OACs giving higher ratings overall than AD, $F(1,30) = 13.730$, $p = .001$, $\eta_p^2 = .314$, with no effect of measure, $F<1$, nor an interaction, $F(1,30) = 1.539$, $p = .224$, $\eta_p^2 = .049$. French words also showed a main effect of group, $F(1,30) = 16.754$, $p < .001$, $\eta_p^2 = .401$, but no effect of measure, $F(1,30) = 0.124$, $p = .728$, $\eta_p^2 = .005$ or interaction, $F(1,30) = 3.680$, $p = .067$, $\eta_p^2 = .0128$.

For nonwords, a main effect of group was present, $F(1,30) = 20.260$, $p < .001$, $\eta_p^2 = .420$ as previously. No main effect of measure was found, $F(1,30) = 0.018$, $p = .894$, $\eta_p^2 = .001$, however a marginal interaction was present, $F(1,30) = 6.355$, $p = .018$, $\eta_p^2 = .185$. Follow up t-tests revealed that the AD group showed no difference between global FOK and actual recognition accuracy, $t(15) = 1.710$, $p = .108$, whereas the OAC group gave marginally lower global FOK ratings compared to their actual recognition accuracy for nonretrieved items, $t(15) = 2.006$, $p = .066$. The AD group thus show an advantage over the OACs, who underestimate the proportion of items they will successfully recognise.

Table 5.12: Summary of main effects and interactions of 2 (group) x 2 (measure) ANOVAs for GFOK within each category

	Fruit	Colours	Creatures	French words	Nonwords
Group	≈	✓	✓	✓	✓
Measure	X	X	X	X	X
Interaction	X	X	X	X	≈

Note: ✓ indicates significant effect, X indicates no significant effect, ≈ indicates marginal effect

In sum, a consistent effect of group was found, with OACs generally higher than the AD group for their GFOK ratings and recognition accuracy in line with actual performance (Table 5.12). Examination of Figure 5.8 would appear to show an underestimation of future recognition in older adults, while the dementia group overestimate recognition in four of the five categories. However, no main effect of measure was observed and interactions also failed to reach significance. Thus, similar to the JORF results, although dementia patients recognised fewer items than the older adult group, they also gave lower predictions about their recognition. It should be noted that a marginal interaction between group and measure was present for the nonword category, especially due to the direction of the interaction. Follow up analyses revealed that the AD group showed no difference between GFOK and recognition, while OACs predicted they would recognise fewer items than they actually were able to at test. Thus OACs showed a level of underconfidence in their GFOKs for nonwords only.

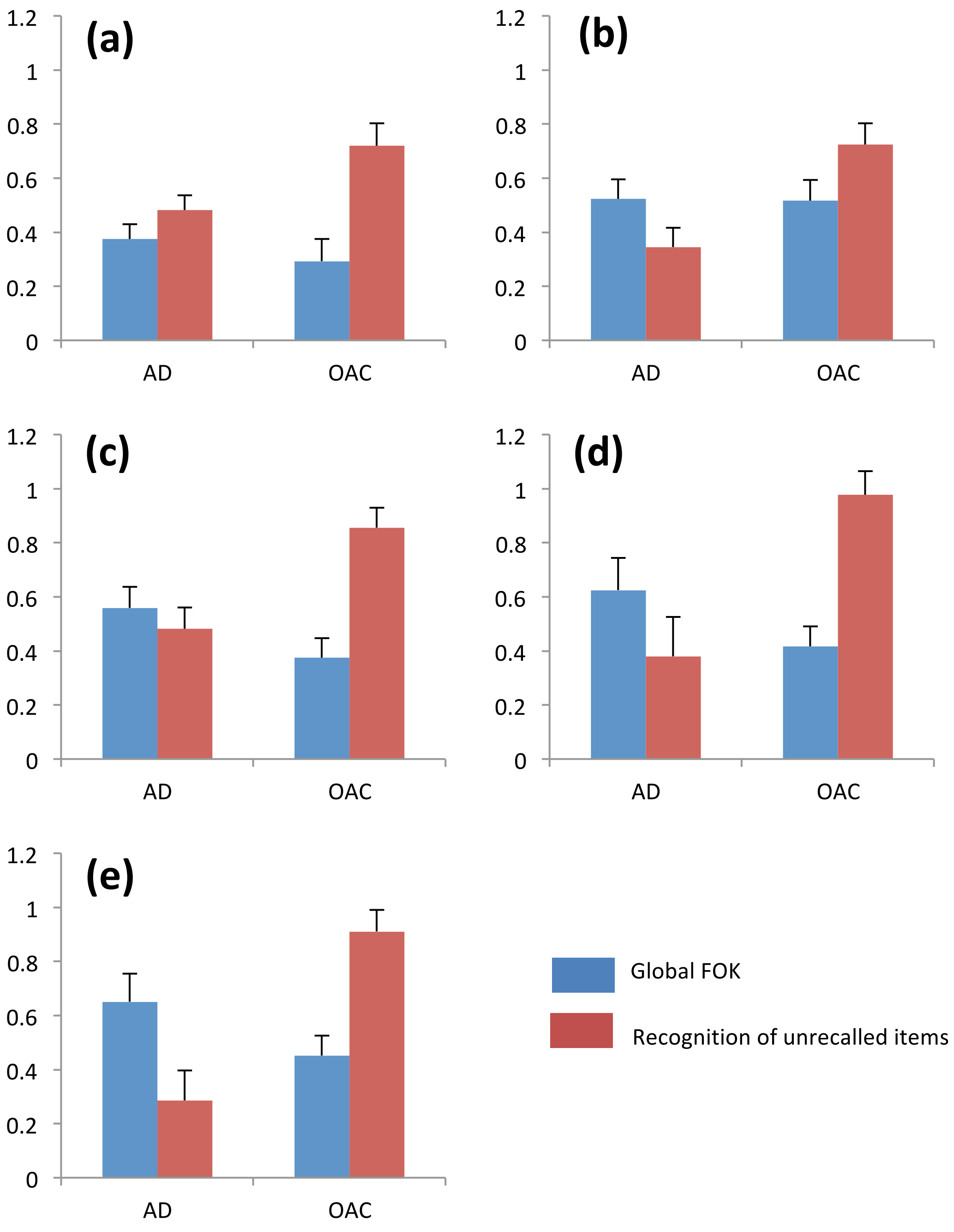


Figure 5.8: Proportion scores for Global FOK and Recognition for Unrecalled Items for (a) Fruits, (b) Colours, (c) Creatures, (d) French words, and (e) Nonwords in the Alzheimer's (AD) group and Older Adult Control (OAC) group. Ordinate represents proportion scores. Bars represent the standard error of the mean.

Table 5.13: Pearson's correlations between GFOK and recognition of unrecalled items for each category within each participant group

	AD		OAC	
	r	*p*	r	*p*
Fruits	.224	*.404*	.110	*.686*
Colours	.547	*.028*	.476	*.062*
Creatures	.753	***<.001****	-.169	*.533*
French words	.595	*.015*	-.116	*.735*
Nonwords	.055	*.839*	.104	*.723*

Note: N = 16 for AD and N = 16 for OAC. * indicates correlation is significant at alpha level of 0.01.

Pearson's correlations revealed that, for the majority of categories, no relationship occurred between GFOK and recognition of unrecalled items (Table 5.13). Only for the creatures category within the AD group could a clear association be observed, with a strong positive correlation between predictions and performance. Thus for this category, as recognition of unrecalled items increased, so too did predictions of future performance. As observed within the JORF category, although no significant difference was observed between the two measures, the GFOK and recognition are not found to be correlated.

5.4.4 Results II: Analysis by list length

Counterbalancing was employed in the current study, with two versions of the materials prepared. It was not possible to assess all category types at all possible set sizes due to the high difficulty level for the French words and Nonwords categories, especially for the patient group. However, it is possible to use the data to provide a preliminary analysis to determine the effect of list length on the memory and metamemory measures obtained. As previously measures of the number of items correctly recalled, discrimination index and JORF were divided by list length, and the global FOK measure was divided by JORF to give a proportion of assumed remaining set that would be recognised.

5.4.4.1 Memory

Recall was examined with a 2 (group: AD and OAC) x 5 (list length: 3 items, 4 items, 5 items, 8 items, and 11 items) mixed ANOVA (see Figure 5.10). As expected, a main effect of group was present, $F(1,30) = 51.975$, $p < .001$, $\eta_p^2 = .634$, with OACs recalling a greater proportion of items than the AD group. No effect of list length was observed, $F(1,30) = 1.373$, $p = .248$, $\eta_p^2 = .044$, nor was there an interaction between group and list length, $F(1,30) = 1.195$, $p = .317$, $\eta_p^2 = .038$. Recognition as measured by the discrimination index was also subjected to a 2 (group) x 5 (list length) mixed ANOVA (see Figure 5.9). Again, a main effect of group was present, $F(1,30) = 62.913$, $p < .001$, $\eta_p^2 = .677$, with OACs showing a higher level of discrimination accuracy than the AD group. A main effect of list length was also present, $F(1,30) = 6.039$, $p < .001$, $\eta_p^2 = .168$. Bonferroni post-hocs confirmed that the discrimination index for list length 8 was significantly lower than that of list length 3 ($p = .015$), and list length 8 and list length 11 were both significantly lower than that for list length 4 ($p = .015$ and $p = .012$ respectively). Thus it would appear that correctly identifying previously encountered items while avoiding distracters becomes more difficult as list length increases beyond 8 items. An interaction between group and list length was also observed, $F(1,30) = 2.849$, $p = .027$, $\eta_p^2 = .087$. One-way ANOVA were conducted separately for each age group to further explore this interaction. For the AD group, no main effect of list length was observed, $F(4,60) = 1.341$, $p = .265$, $\eta_p^2 = .082$. In contrast, for the OAC group, a main effect of list length did occur, $F(4,60) = 12.769$, $p < .001$, $\eta_p^2 = .460$. Bonferroni post hocs confirmed that the longer list lengths of 8 and 11 items showed a lower discrimination index than those of list length 3 ($p <.001$ and $p < .001$) and list length 4 ($p = .003$ and $p < .001$).

Table 5.14: Mean (*SD*) intrusion errors made by AD patients and Older Adult Controls at each list length

	3 items	4 items	5 items	8 items	11 items
AD	0.125 (0.342)	0.125 (0.342)	1.313 (2.869)	1.438 (1.413)	1.000 (1.366)
OAC	0.125 (0.342)	0.250 (0.447)	0.625 (1.088)	0.500 (0.632)	0.625 (0.719)

Intrusion errors were also examined with a 2 (group) x 5 (list length) ANOVA (see Table 5.14). As previously, raw numbers of errors were examined rather than using a proportion of list length. No effect of group was found, $F(1,30) = 2.469$, $p = .127$, $\eta_p^2 = .076$. A main effect of list length was present, $F(1,30) = 4.475$, $p = .017$, $\eta_p^2 = .130$, with the number of intrusion errors increasing as list length increased. No interaction was present, $F(1,30) = 1.264$, $p = .289$, $\eta_p^2 = .040$.

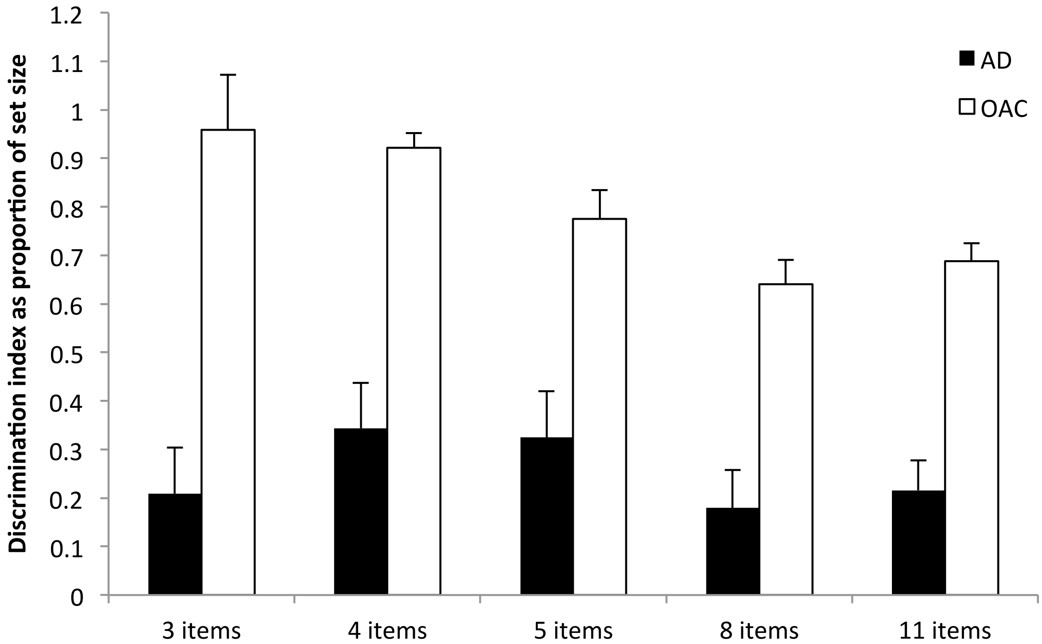


Figure 5.9: Discrimination index (hits-false alarms) as a proportion of category size for Alzheimer's Disease (AD) patients and older adult controls (OAC)

5.4.4.2 JORF

The accuracy of the JORF was examined using a 2 (group) x 2 (measure: JORF and actual retrieval failure) x 5 (list length) mixed ANOVA (see Figure 5.10). As for the category-led analysis, a main effect of group was observed, $F(1,30) = 55.969$, $p < .001$, $\eta_p^2 = .651$, with AD patients giving responses of greater magnitude than those of OACs. No main effect of measure was found, $F(1,30) = 1.370$, $p = .251$, $\eta_p^2 = .044$, indicating that JORFs and actual retrieval failure were at similar levels across groups and list length. The effect of list length was marginal, $F(4,120) = 2.867$, $p = .055$, $\eta_p^2 = .087$. Bonferroni post hocs indicated that this effect was driven by list length 11 being significantly lower than list lengths 4 and 5 ($p = .006$ and $p = .035$

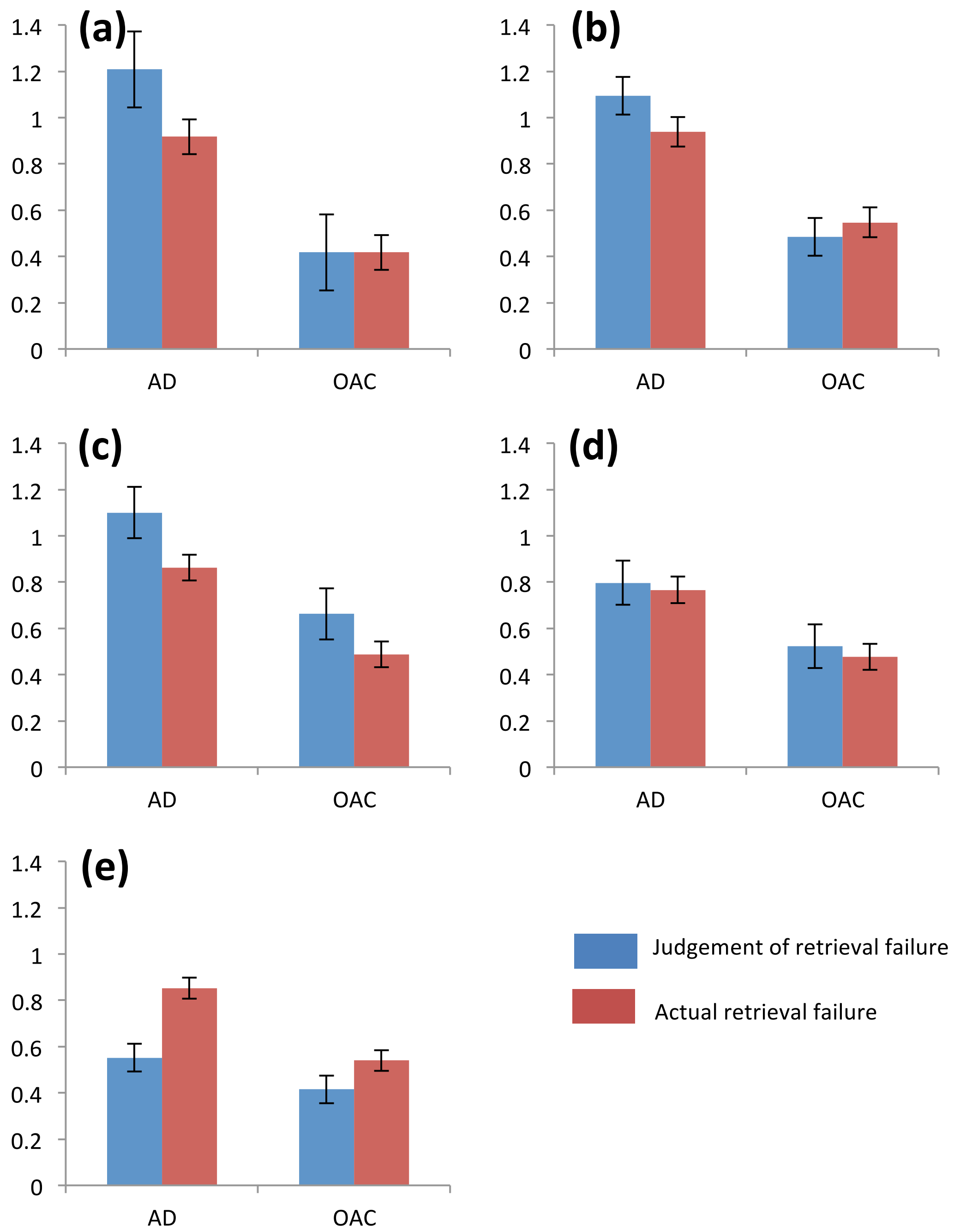


Figure 5.10: Proportion scores for Judgement of Retrieval Failure (JORF) and Actual Retrieval Failure for list lengths of (a) 3 items, (b) 4 items, (c) 5 items, (d) 8 items, and (e) 11 items in the Alzheimer's (AD) group and Older Adult Control (OAC) group. Ordinate represents proportion score

respectively). No interaction was present between group and measure, $F<1$, however an interaction did occur between group and list length, $F(4,120) = 2.971$, $p = .049$, $\eta_p^2 = .090$. While OACs respond in a fairly stable manner over the different list lengths, AD patients show a higher magnitude of responses for shorter list lengths and a lower magnitude of responses at the longest list lengths of 8 and 11 items. An interaction between list length and measure was also found, $F(4,120) = 6.617$, $p = .003$, $\eta_p^2 = .181$. While list length appeared to have no effect on the proportion of items which were failed to be retrieved, $F(4,124) = 1.364$, $p = .259$, $\eta_p^2 = .042$, the JORF showed a steady decline over list length, $F(4,124) = 4.322$, $p = .016$, $\eta_p^2 = .122$, with a significantly lower proportion of items judged to be remaining at list length 11 than for list lengths of 4, 5 and 8 items (Bonferroni posthocs: $p = .002$, $p < .001$ and $p = .040$ respectively). No interaction was present between group, measure and list length, $F(4,120) = 2.237$, $p = .117$, $\eta_p^2 = .069$.

5.4.4.3 Global FOK

GFOKs were also examined with a 2 (group) x 2 (measure: GFOK and recognition of nonretrieved items) x 5 (list length) mixed ANOVA (see Figure 5.11). A main effect of group was present, $F(1,23) = 20.575$, $p < .001$, $\eta_p^2 = .472$. In contrast to the JORF findings, for GFOK OACs gave responses of greater magnitude than did the AD group. Main effects of measure and length both failed to reach significance, $F<1$. An interaction was observed between group and measure, $F(1,23) = 6.511$, $p = .018$, $\eta_p^2 = .221$. While the AD group gave higher GFOK than actual recognition of nonretrieved items, the OAC group gave slightly lower GFOKs compared to their recognition of nonretrieved items. Thus while the AD group show a level of overconfidence in their recognition ability, OACs tend to underestimate their ability. Interactions between group and list length, $F(4,92) = 1.117$, $p = .346$, $\eta_p^2 = .046$, measure and list length, $F(1,23) = 1.495$, $p = .217$, $\eta_p^2 = .061$, and between group, measure and list length, $F(4,92) = 1.695$, $p = .167$, $\eta_p^2 = .069$, all failed to reach significance.

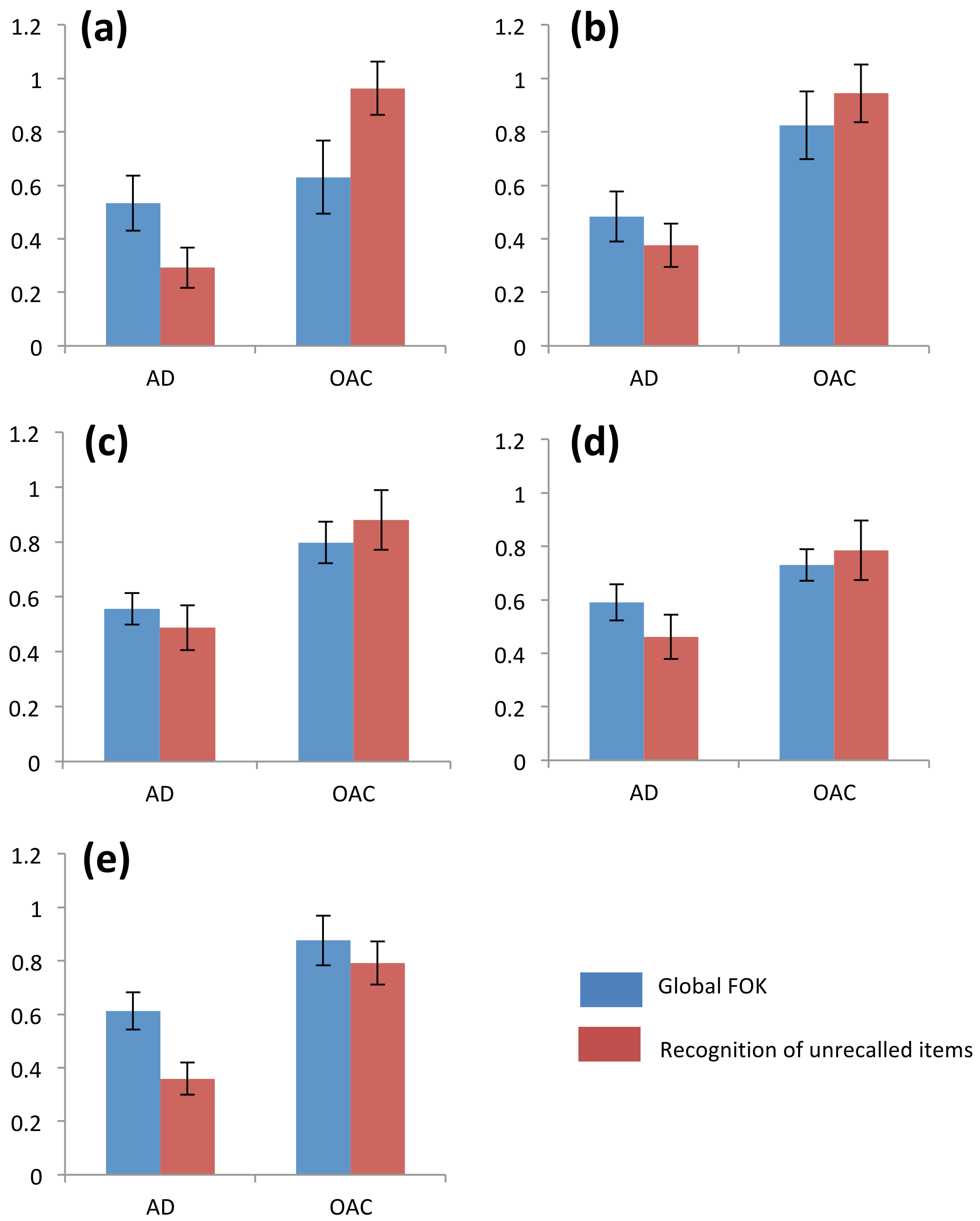


Figure 5.11: Proportion scores for Global FOK and Recognition for Unrecalled Items for list lengths of (a) 3 items, (b) 4 items, (c) 5 items, (d) 8 items, and (e) 11 items in the Alzheimer's (AD) group and Older Adult Control (OAC) group. Ordinate represents proportion score

5.4.5 Discussion

The aim of the current experiment was to explore JORFs and GFOKs within a memory impaired patient sample. The JORF offers a novel measure of memory awareness, while problems obtaining episodic item-by-item FOKs could be avoided by employing a global judgement of future recognition ability for unrecalled items. As expected, patients diagnosed with dementia showed a memory deficit in comparison to healthy older adult controls, with lower levels of recall and recognition accuracy. Both healthy older adults and dementia patients were able to make the metacognitive judgements without difficulty.

The JORF requires monitoring of memory failures, the ability to judge how much of a block of information has not been retrieved. In Experiment 5.1 with young adults, there was a general tendency to overestimate the number of remaining items from the category in question, with three out of five categories exhibiting this pattern. In the present study, participants gave JORFs which accurately reflected the remaining items in the category. Neither patients with dementia nor healthy older adults showed evidence of under- or overestimation in their JORF predictions. This raises the interesting possibility that older adults and patients are actually better able to judge this type of memory sensation than young adults. Further exploration of this with matched materials and difficulty level would be required to confirm this suggestion.

The similarity between JORF ratings and retrieval failure for the patient group is somewhat surprising in light of the well documented deficits in metacognition typically shown within this group (Pannu & Kaszniak, 2005; Souchay, 2007). For global judgements especially, a number of studies have found evidence of overestimation of ability in confidence judgements and JOLs (e.g. Barrett et al., 2005; Schacter et al., 1986). Semantic processes appear to be preserved in early AD, with evidence from both direct examination of monitoring abilities for semantic memory (Bäckman & Lipinska, 1993; Lipinska & Backman, 1996) and their ability to use knowledge of general memory effects to adjust metacognitive judgements (Moulin et al., 2000c). It may therefore be that these preserved semantic memory processes may be recruited to aid accuracy. However, the findings of Experiment 5.2 would suggest against this as a potential explanation. Here, judgements for the semantic task showed poor resolution, with participants underestimating the number of unrecalled items and no relationship being observed between JORFs and actual retrieval failure. It would therefore be unlikely that recruitment of semantic processes would allow judgements within the current study and current participant groups to become more accurate.

Results from the GFOK measure are less conclusive as to whether a deficit in monitoring occurs in AD patients. Previous examination of item-by-item FOKs suggest that AD patients are unable to predict future recognition of unrecalled items, with evidence of overconfidence in their future memory ability (Souchay et al., 2002). Two categories within the current study, colours and French words, also showed this pattern, with AD patients predicting they would recognise a greater proportion of unrecalled items than they were able to. Healthy controls were, in contrast, accurate in their metacognitive assessment. However, the nonword category showed AD patients to be accurate in their GFOK judgements while healthy OACs underestimated the amount of information they would later recognise, while for creatures and fruits both participant groups were equally accurate on their judgements. Thus older adults are relatively consistent in their GFOK accuracy, while the accuracy of AD patients deviates depending on the category. This variability in the accuracy of AD patients may be indicative of a general inability to accurately predict future recognition of unrecalled items. It is reasonable to assume that some judgements may reflect actual performance by chance. For example, in Souchay et al.'s (2002) study it may be that a subset of the nonrecalled items within the AD group showed good predictive accuracy, whereas another subset may not. On the whole however, a deficit in metacognitive accuracy would remain. In the current study only five judgements are made, therefore without a consistent pattern of results it may be assumed that patients have difficulty making an accurate GFOK judgement.

Although counterbalancing was incomplete, the use of varied item numbers for categories enabled a preliminary analysis of the effects of list length on the memory and metacognitive measures obtained. While no effects were found for recall, recognition accuracy did decrease as list length increased. Similarly, an increase in intrusion errors was observed with increasing list length. With a larger possible set of targets available, participants were more likely to incorrectly identify category exemplars as targets and report them during recall and recognition tests. This replicates previous findings examining varying list lengths in the memory literature (e.g. Cary, 2003). Additionally, JORF was found to be affected by list length. With participants unable to accurately identify which category exemplars were present at encoding, it would be logical to assume that they would also have difficulty judging the number of items they had failed to recall. Beyond a list length of around 8, the category becomes too large to have an accurate concept of the precise number, whereas for smaller categories the range of possible error is much smaller and so judgements, and memory, are more accurate.

The present study has established the JORF and GFOK measures as future avenues of metacognitive investigation in AD and dementia. In agreement with previous research suggesting aspects of metacognition may be preserved in early dementia, patients within the current study were able to accurately monitor the amount of information they had failed to recall. Memory accuracy was lower in the dementia group than for healthy older adults, yet JORFs were higher, thus patients were able to adjust their judgements to reflect performance. In contrast, patients were unable to make GFOK judgements, with inconsistent accuracy throughout the categories presented. This suggests that metacognitive deficits within dementia and AD are not universal, and some aspects of memory monitoring may be preserved in the early stages of the disease.

5.5 General discussion

The current chapter utilised a new metacognitive measure, the judgement of retrieval failure (JORF) and incorporated a novel global FOK measure (GFOK). To date, metacognitive measures have not examined global assessment of failures of memory. The standard FOK paradigm does allow some examination of people's ability to assess the unrecalled contents of memory. However, this is on an item-by-item basis which is not necessarily representative of how everyday memory monitoring occurs. In addition, the FOK requires a prompt to be given to the participant to elicit a retrieval attempt of the item. This provides the learner with the knowledge that a particular item should be in memory and therefore make a judgement as to whether it will be retrieved later. The JORF removes this prompt, so the learner is more directly assessing whether the item is in memory. If a particular item has not been encoded to a sufficient extent, no memory trace will have formed; therefore it will not be included in the global assessment of retrieval failure.

Over three experiments, it has been shown that participants are able to make episodic JORFs for retrieval failures from various semantic categories. Experiment 5.1 demonstrated that the JORF is sensitive to the classic memory manipulation of delay, with JORFs increasing as retention interval (and thus retrieval failure) increased. Moderate correlations were also observed between immediate JORFs and retrieval failure, again indicating their sensitivity to memory. Experiment 5.2 provided a preliminary investigation of the accuracy of JORFs for episodic and semantic materials. Findings indicated that semantic JORFs were less accurate than episodic JORFs, although conclusions are tentative due to differences in set seize which may have affected the ability of participants to make judgements for the semantic materials. Experiment 5.3 showed the JORF to be a useful tool in evaluating metacognitive abilities in a patient population. Participants diagnosed with AD were found to be able to adjust their JORF assessments to reflect their increased failure to recall items. The JORF is therefore an informative measure to add to the metacognitive arsenal.

The level of counterbalancing of set size used within the current chapter was necessarily limited due to the participants and testing environments employed. Experiment 5.1 occurred in a lecture setting, thus only one version of the task could be presented. Likewise, although some counterbalancing was established in Experiment 5.3, due to limitations in participant ability two of the three categories required a lower set size to keep participants engaged with the task. The patterns of under- and overconfidence observed within the current findings could therefore be either due to the set size used, or due to the semantic category. It could be argued that the categories themselves may be having an influence, especially as they do not equally rely on semantic and episodic memory processes. While fruits, creatures, colours and even French words all have some semantic representation in memory which may be influencing judgements, the nonwords have no such representation to influence the JORF and GFOK, which may consequently affect their magnitude and resolution. Differences in set size may appear to be a more viable route of explanation for the differences in under-and overconfidence. Indeed, Tauber & Rhodes (2010) have already noted that the amount of material to be remembered can have an influence on JOLs. Although JOLs were found to be sensitive to list length, they also became less accurate as list length increased.

Data from the current chapter are not able to answer whether changes in JORF accuracy are due to category type or due to list length. Preliminary analysis of list length effects in Experiment 5.3 indicates that JORF decreases as list length increases while retrieval failure remains constant. This would result in a tendency to underestimate retrieval failure at longer list lengths. If we exclude the nonwords category, this is indeed what is found in Experiment 5.1, where French words (7 items), Colours (10 items) and Fruits (13 items) all show overestimation while Creatures (17 items) shows underestimation. However, JORFs for nonwords (9 items) were not found to be significantly different to retrieval failure, which would not fit the pattern suggested. Thus it may be that both category type *and* list length affect JORF accuracy, with JORFs decreasing with increasing list length when semantic categories are used, and being unaffected by list length when semantic categories are not used. This can easily be dissociated through further investigation within settings which allow for a fully counterbalanced design.

Throughout the chapter the argument is made that a non-significant difference between JORF and retrieval failure, and between GFOK and recognition of unrecalled items, is an indication of accuracy. This mirrors the argument used in global JOLs whereby discrepancy scores, the difference between judgement and performance, are used to infer accuracy (e.g. Correa, Graves, & Costa, 1996). This argument is supported by the presence of a moderate positive correlation between the measures within Experiments 5.1 and 5.2, where the large sample sizes used allow greater confidence that that random factors are not responsible and the power of the analysis is high. Experiment 5.3 does not however provide convincing correlational evidence for a relationship between the measures. There are two possible reasons for this. First, the sample size is small, thus the presence of a correlation is more difficult to detect. Second, it may be that the participants used do not make their predictions in a similar way to the young adults within the other experiments. Patients with dementia, although sensitive to some task effects, nonetheless exhibit deficits in metacognitive accuracy (Souchay & Moulin, 2009; Souchay, 2007). Older adults have also been found to show deficits in metacognitive monitoring, as demonstrated in Chapter 2. Subsequently, correlations between the measures may not be present due to difference between the information used when making judgements by young adults and by older adults and patients with dementia. Indeed, studies within the FOK literature have indicated that older adults do not spontaneously use information to inform their predictions in the way young adults do (Thomas et al., 2011). Nonetheless, the use of a non-significant difference to infer accuracy should be approached with caution.

It is worth considering the potential presence of a mid-point anchoring effect observed in other global metacognitive judgements (Connor, Dunlosky, & Hertzog, 1997). Participants have a general idea of their memory ability for a variety of tasks; this forms a part of our knowledge of how our own memory works. When presented with a novel memory task, as is typical in experimental investigations in metacognition, participants have a tendency to 'anchor' their judgements at the midpoint of this known performance range. Their actual memory accuracy may be different to this midpoint, due to normal variations in memory dependent on the task characteristics, thus determining the accuracy of the metacognitive judgement. Although not the sole contributor, mid-point anchoring does nonetheless make an important contribution to the magnitude of JOLs given, providing a preliminary basis which is then amended according to information provided by monitoring processes (Scheck, Meeter, & Nelson, 2004).

The impact of this effect within the JORF is difficult to ascertain. Unlike the JOL, the JORF paradigm does not provide a scale on which to make a judgement, thus participants cannot use a mid-point as the basis of their judgement. It is however reasonable to suggest that participants may have a general concept of their ability to recall lists of varying lengths which may be influencing their assessment of the amount of information they have failed to recall. Consideration of the means in Experiment 5.1 does not appear to show a mid-point effect. However, it is interesting to note that for the categories containing 10 or fewer items mean JORFs are a large proportion of category size (colours, 0.854; French words, 0.926; nonwords, 0.840), while the categories above 10 items are around 50% of category size (fruits, 0.519; creatures, 0.594). With adequate counterbalancing this potential influence on JORF accuracy could easily be examined in future studies to ascertain whether mid-point anchoring can occur when no referent for performance is given.

A further avenue of extended research with the JORF paradigm is the effect of framing. Framing has been extensively researched within the area of decision making (for reviews see Kühberger, 1998; Levin, Schneider, & Gaeth, 1998), with the finding that the framing of a statement or question can influence how people make judgements and decisions. The type of framing of particular relevance to the metacognitive literature is termed attribute framing (Levin et al., 1998). Here, only a single attribute of the given context is changed to alter the framing, for example evaluating an item in terms of whether it is good or bad. The alteration of the valence of the descriptive statement influences the processing of that statement, thus affecting the judgement (or decision) that is made. Koriat, Bjork, Sheffer, & Bar (2004), as discussed in Chapter 3, demonstrated the effect of framing a question in terms of forgetting rather than remembering for different retention intervals. When asked to judge recall accuracy following various delays, participants showed no variation in predictions. However, when asked to judge forgetting over the same delays, participants increased their judgements (thus reducing recall predictions) as delay increased. The framing caused participants to consider different cues in order to make their judgements, affecting their accuracy. This effect has also be examined with immediate JOLs (Finn, 2008; Kornell & Bjork, 2009), with judgements of forgetting the target word in a pair closer to actual recall than judgements of remembering (but see also Serra & England, 2012). Previous research has already considered the accuracy of global judgements of recall accuracy, with assessments either accurately reflecting performance or showing a slight overestimation (Bunnell et al., 1999; Hertzog et al., 1994). The JORF could be considered the alternative framing of this recall judgement; presumably retrieval failure would be equivalent to 1 minus the proportion judged to be recalled. It would be of interest to consider the accuracy of JORFs and judgements of recall accuracy within the same experimental task to establish if this is indeed the case, or if the framing of the question in terms of judging recall or retrieval failure may affect the accuracy of said judgement.

There are a number of further considerations with the JORF that may be interesting to explore. There is an ambiguity in Experiment 5.2 as to whether effects were due to the material used (i.e. semantic versus episodic) or due to the set sizes of the materials. The semantic task used may have been too difficult for participants to be able to judge actual retrieval failure. Without some level of knowledge of the periodic table, it would not be possible to judge the remaining unrecalled items. Although all participants will have encountered the periodic table during school, they may not have had to learn the individual elements, thus would not have been able to use any partial information relating to unrecalled elements to make their JORF assessments. It would also be interesting to ask participants whether they were aware of the different semantic categories used for the episodic task in Experiment 5.2, as this may have influenced strategies at encoding and when making the JORF prediction. The influence of visual memory also needs consideration. Studies into visual search have ascertained that if the same configuration of items is used on search trials, detection of a target item is enhanced (Thomson & Milliken, 2013; Travis, Mattingley, & Dux, 2013). This suggests that the spatial relationship between items is represented within the memory system, allowing for contextual cueing to aid target recognition. It is therefore possible that a spatial representation was also formed within the encoding of materials in the current tasks. During the retrieval attempt, participants may have spontaneously retrieved the configuration of items as well, perceiving the encoding context within the “mind’s eye”. This representation may subsequently contribute to the partial information on which the metacognitive judgements were made.

Throughout the chapter, it has been assumed that the JORF is based on partial information about the individual items which have not been recalled. This implies that a JORF of 6 means that 6 items are associated with enough partial information for the participant to believe that they were present during encoding. The JORF would therefore be a one-stage judgement; monitor partial information and produce the judgement. However, the JORF may also be considered as a two-stage problem solving task. Participants may use knowledge of their recall performance and knowledge of the target set size to make the JORF, thus subtraction of the number of items recalled from the set size would give an assessment of retrieval failure. Indeed, this route was considered when choosing the semantic task for Experiment 5.2, whereby it was ensured that the target set size was not known as an independent fact from the items themselves (the example of Snow White’s seven dwarfs). Thus the JORF is not based on partial information associated with individual items, but is based on knowledge of recall and set size. The question then arises as to how participants would know the set size for episodic materials. It may be that participants counted the number of items presented. Results would suggest this is not the case, as if participants were routinely counting the number of items presented JORFs would be consistently matched to actual retrieval failure. This could easily be verified by asking participants to report what strategy was used during encoding. Alternatively, participants may have a belief about their recall ability. Similar to the mid-point anchoring effect, participants may believe that they reliably recall 50% of a list, and so judge the remaining number of items based on the amount they have successfully recalled. As previously noted, there does appear to be a tendency to for JORFs to be around 85% of retrieval failure for set sizes below 10 and around 50% for set sizes above 10. This explanation does therefore warrant further exploration, both with regards to the basis of JORF and with regards to the influence of set size on metacognitive judgements. Finally, it may be that the ease with which items are recalled may be used as an indicator of set size. For some semantic categories, it is relatively easy to produce category exemplars, whereas other categories may be more difficult, e.g. fruits as compared to modes of transport. Eliciting descriptions of any strategy use when making metacognitive judgments may shed further light as to whether participants rely on partial information from individual items or use a problem-solving methodology to give their JORFs.

The GFOK is a second global metacognitive measure which the present chapter has established as a potential avenue for future research. Within Experiments 5.1 and 5.3 it has been used in conjunction with the JORF, however it may also be considered as a stand-alone measure of monitoring if participants were prompted with the category size. As with the JORF, this may lead to a mid-point anchoring effect, with GFOKs reflecting a consistent proportion of unrecalled items rather than a true estimation of future recognition ability. It may be that increased experience with a particular task would allow participants to more accurately make global judgements of recognition of unrecalled items, thus updating their knowledge of how their memory performs. As in the JORF, the effect of semantic categories and differing set sizes needs clarification. Although the present studies examine all measures in terms of proportions, this cannot fully account for the differences in category sizes and items presented at encoding. It would therefore be beneficial to also consider GFOKs in the suggested future research of the JORFs.

The present chapter has successfully established two novel measures of global metacognitive judgements. In addition the two judgements, while informative as individual measures, also complement each other. The JORF is a retrospective assessment of retrieval failure, while the GFOK is a prospective judgement of memory accuracy in light of this retrieval failure. The assumption may therefore be that similar information is used when making these judgements, an assumption that is beyond the scope of the present chapter to answer but is an interesting possibility to explore in future studies. The judgements have also been shown to be suitable measures of metacognition for a memory impaired sample to use, raising the prospect of more novel insights into the monitoring abilities of other patient groups. On a broader level, it may also be interesting to consider how these judgements may be used by participants to alter their learning behaviour. Our capacity to monitor our memory processes is fundamental to maximising our memory ability. The JORF and GFOK measures may shed further light on how this assessment of memory sensations enables us to adjust encoding strategies to aid recallability.

Chapter 6. General Discussion

6.1 Overview

This thesis aimed to explore the relationship between memory and evaluations of retrieval failure. One major theory underlying metacognitive research is that memory function is inextricably linked with the production of an FOK (Perfect & Stollery, 1993). Thus, the overarching goal was to examine how factors which influence memory affect the magnitude and accuracy of feeling of knowing predictions. This was achieved through experimental manipulations of memory at encoding, retention and retrieval, and also through consideration of the impact of healthy ageing in conjunction with these task effects. A second aim of this thesis was to develop two novel metacognitive measures of retrieval failure, and to establish the impact of manipulations of memory on these measures. In addition to experimental manipulations and healthy ageing effects, the impact of early stage dementia on these metacognitive abilities was also investigated. The present chapter summarises the key findings of the thesis and considers the wider theoretical implications of these findings. This is followed by discussion of methodological limitations of the present work and metacognitive research generally, and highlights some future directions for research to extend on the research presented here.

6.2 Summary of findings

The primary measure used throughout the thesis was the FOK paradigm. Table 6.1 presents a summary of the experimental manipulations which were considered in Chapters 2 and 3, and their effects on recall accuracy, the magnitude of FOK judgements given (or the proportion of Yes FOK responses given) and the subsequent changes in FOK accuracy as measured by gamma. The effect of ageing on recall and FOK was also examined throughout Chapter 2 and 3. Table 6.2

presents a summary of the differences in recall, FOK magnitude and FOK accuracy observed within each condition of these experiments.

Table 6.1: Summary of experimental manipulations and their effects on recall and FOK

	Manipulation	Effect on recall	Effect on FOK magnitude	Effect on FOK accuracy
Exp. 2.1	*Semantic (S) vs Episodic (E)*	S < E	S < E	S = E *
Exp. 2.2	*Semantic (S) vs Episodic (E)*	S > E	S < E	S > E
Exp. 2.2	*Learning*	Increased	Increased	Increased
Exp. 3.1	*Increased Delay – between subjects*	Decreased	Decreased	No effect
Exp. 3.2	*Increased Delay – within subjects*	Decreased	Decreased	No effect *
Exp. 3.3	*Increased Environmental support*	Increased	Increased	No effect
Exp. 3.4	*Increased Environmental support*	Increased	Increased	Increased

* indicates interaction with age

Table 6.2: Summary of age effects on recall and FOK

		Effect on recall	Effect on FOK magnitude	Effect on FOK accuracy
Exp. 2.1	*Semantic task*	OA > YA	OA > YA	OA = YA
	Episodic task	OA = YA	OA = YA	OA < YA
Exp. 2.2	*Semantic task*	OA > YA	OA = YA	OA = YA
	Episodic task	OA = YA	OA = YA	OA < YA
Exp. 2.2	*Learning*	OA = YA	OA = YA	OA < YA
Exp. 3.2	*Immediate*	OA < YA	OA = YA	OA = YA
	Delayed	OA < YA	OA = YA	OA > YA
Exp. 3.3	*2 cues*	OA = YA	OA < YA	OA = YA
	4 cues	OA = YA	OA < YA	OA = YA
Exp. 3.4	*2 cues*	OA < YA	OA < YA	OA = YA
	4 cues	OA < YA	OA < YA	OA = YA

OA = Older adults, YA = Young adults

6.2.1 Semantic and episodic FOK in ageing

Semantic and episodic memory processes differ in both the type of information involved and the subjective experiences associated with them (Squire, 1984; Tulving, 1972, 1985b). Healthy ageing has differential effects on each type of memory, with semantic processes preserved while episodic processes are impaired (Anderson & Craik, 2000; Craik et al., 1995; Zacks et al., 2000). Examination of the FOKs related to each type of memory and the effect of ageing thus provides a useful tool to consider the link between memory and FOK; do selective deficits in episodic memory in older adults relate to selective deficits in episodic FOKs. Previous findings support both the assumption of a selective deficit in episodic FOK accuracy (Souchay et al., 2000, 2007; Thomas et al., 2011), and preservation of episodic FOK accuracy (MacLaverty & Hertzog, 2009). The novel paradigm used within Chapter 2 allowed examination of both semantic and episodic FOK accuracy within the same task and the same participants, thus minimising the influence of task and item characteristics.

With general knowledge items, Experiment 2.1 established preserved semantic FOK accuracy within older adults, as expected. In contrast, episodic FOK judgements were found to be impaired, with older adults at chance accuracy and significantly lower than young adults. This was despite episodic recall performance being significantly higher in the older adult group. Thus the episodic FOK deficit occurred in the absence of an episodic memory deficit. Following a similar paradigm with French-English translations, Experiment 2.2 again found preserved semantic FOK judgements for older adults; predictive accuracy was similar to that observed in young adults. Although no main effect of age was found for episodic FOK accuracy, older adults were found to be at chance level of performance, unlike young adults who were reliably above chance. This absence of predictive accuracy within the older adult group only could therefore indicate a selective deficit in episodic FOK predictions, in agreement with the findings of Experiment 2.1. Furthermore, no difference was found between young and older adults for episodic memory accuracy. Deficits in FOK accuracy again occurred in the absence of deficits in episodic memory.

The findings of Chapter 2 suggest a dissociation between memory and FOK accuracy dependent on the type of memory examined. While semantic memory and its associated FOK judgements were found to be similar in young and older adults, episodic memory was not necessarily associated with episodic FOK. Older adults exhibited deficits in episodic FOK accuracy despite equivalent or superior episodic memory. Findings therefore go against the assertion of the memory constraint hypothesis (MacLaverty & Hertzog, 2009) that previously observed deficits in episodic FOK accuracy in ageing are due to age-related deficits in encoding processes. Rather, the observed deficits in episodic FOK accuracy appear to reflect a true deficit within the metacognitive processes which form the FOK judgement in ageing (Souchay et al., 2000, 2007; Thomas et al., 2011).

In addition to exploring how memory type affects the FOK, Experiment 2.2 also examined whether increasing memory accuracy would lead to increases in FOK accuracy (Carroll & Nelson, 1993; Nelson et al., 1982). With repeated learning trials, both young and older adults were able to not only increase the proportion of items correctly recalled but also to increase the predictive accuracy of the FOK judgements given for unrecalled items. Despite the improvements in FOK accuracy observed, older adults did remain at chance level of accuracy until the final episodic trial; although older adults improved throughout the task they were unable to reach the levels of accuracy obtained by the young adult group. These findings do however raise the possibility that older adults may be able to take advantage of an intervention or training to reduce the impact of age-associated metacognitive deficits.

The findings of Chapter 2 suggest that, for young adults, manipulations at encoding via repeated learning trials and manipulations of memory type lead to equivalent effects on memory and FOK. For older adults, while the manipulation of encoding resulted in equivalent memory and FOK effects, the change of memory type did not. Specifically, episodic FOK was impaired while episodic memory was not. The question then arises as to why this discrepancy in memory and FOK exists within this population. Souchay et al. (2007) suggest that older adults have difficulty making episodic FOK judgements due to associated deficits in recollective experience. Indeed, higher FOK judgements have been shown to be associated with higher recollective experience (Hicks & Marsh, 2002). However, Experiment 2.1 failed to find any such association between episodic FOK and the level of Remember responses given by either young or older adults, suggesting this cannot be the sole contributor to the observed age effect.

Another potential avenue of interest is the role of executive function (Perrotin et al., 2008; Souchay et al., 2000; Souchay & Isingrini, 2004). Executive control and FOK are conceptually similar, with both involving the monitoring of performance, setting goals, and changing behaviour dependent on feedback (Fernandez-Duque et al., 2000). Furthermore, executive function is dependent on the frontal lobes, an area which is particularly vulnerable to the effects of ageing (Raz et al., 2005), and which has also been implicated in episodic FOKs (Reggev et al., 2011). Because of these observed similarities, it is important to consider again whether the observed FOK deficit is due to deficits in the metacognitive processes themselves, or if it is due to deficits in related cognitive processes.

6.2.2 Retention and retrieval

Further examination of the memory-FOK link was established through experimental manipulation of retention and retrieval conditions in Chapter 3. A two hour delay was introduced at varying stages of the FOK procedure in Experiments 3.1 and 3.2, leading to lower recall accuracy in young and older adults. While the magnitude of FOKs mirrored changes in recall performance, no effect on the accuracy of the FOK judgements given was observed. The level of environmental support provided at retrieval in Experiments 3.3 and 3.4 had similar findings. While increased environmental support led to higher recall accuracy and higher FOK predications in young and older adults, effects for FOK accuracy were inconclusive. Experiment 3.4 suggested increased accuracy in line with increased environmental support, while no effect was observed in Experiment 3.3. Thus FOK magnitude is sensitive to memory accuracy following manipulations at retention and retrieval, whereas FOK accuracy appears for the most part unaffected by these changes in memory performance.

Experiment 3.1 varied the placement of a two hour delay within the FOK procedure and subsequently the information on which the FOK was based. By placing the delay after the FOK prediction but before recognition, FOK judgements were lower in magnitude than when no delay was present. In this condition, participants are required to not only predict their future recognition of unrecalled items but also predict the effect of the two hour delay on this future recognition. Thus in addition to being experience-based, i.e. based on the partial information retrieved, the FOK is also theory-based, i.e. taking into account the known effects of a delay on memory performance (Koriat et al., 2004). This reliance on theory-based information results in the reduced FOK magnitude. When the delay is placed before both the FOK and recognition, participants can rely on experience-based judgements only, as the delay has already affected memory and thus the partial information available is diagnostic for the recognition task which will immediately follow. In this condition, participants made FOK judgements which lay slightly lower than when no delay occurred, and slightly higher than when the delay occurred between FOK and recognition. Although not significant, this could indicate that people over accommodate for the effects of delay when incorporating theory-based effects into their FOK judgements.

Experiment 3.2 furthered the investigation into retention interval by eliciting FOKs both before and after the two hour delay. In addition, both young and older adults were tested to establish if ageing interacted with the effects of delay. Recall accuracy was affected by delay and age as expected, with memory lower following the delay and older adults recalling fewer items than young adults. Although no main effects were found for FOK accuracy, an interaction did occur between age and delay, with older adults becoming more accurate following the delay than prior to the delay. The mechanism behind this intriguing result cannot be fully answered with the current paradigm. It may arise due to older adults not incorporating theory-based information after the delay, inadvertently allowing them to give more accurate predictions than young adults, however this remains speculative. With regards to magnitude, FOK ratings were lower after the delay than before, replicating the results of Experiment 3.1.

Chapter 3 also considered the effect of environmental support at retrieval for young and older adults (Craik 1983, 1986). By asking participants to encode multiple items as a single unit, it was possible to vary the number of cues (either two or four) used at retrieval to prompt recall of a specific target. Over two experiments, consistent effects of environmental support on recall accuracy were shown, with young and older adults equally benefitting from a higher number of cues. In addition, the magnitude of FOK ratings was also sensitive to the manipulation of support. FOK ratings for both age groups were higher when four cues were given for the missing target than when two cues were given. With regards to FOK accuracy, gamma correlations were typically at chance for both young and older adults, although older adults did show above chance performance in Experiment 3.4. Additionally, in Experiment 3.4 the manipulation of cues affected FOK accuracy. With more cues available, and thus greater environmental support, FOK predictions became more accurate.

The key finding from Chapter 3 is the evidence of sensitivity in FOK judgements. The magnitude of the FOK ratings provided by both young and older adults were consistently affected by the manipulations of retention interval and environmental support, with a delay leading to reduced FOK ratings while increased support led to increased FOK ratings. Although this did not ultimately have corresponding effects on FOK accuracy, the changes in magnitude do indicate that the manipulations altered the information on which the FOK was based. The addition of an ageing component into the experiments also allowed further consideration of the potential age-associated deficits in episodic FOK accuracy. Interestingly, findings from the current studies were not necessarily in agreement with the findings of Chapter 2. For Experiment 3.2, older adults became more accurate than young adults following a two hour delay. As outlined in the chapter, this could either reflect a true increase in accuracy within the older adult group, or reflect a deficit in the application of theory-based information. For Experiment 3.4, while young adults are at chance accuracy in both support conditions, older adults are above chance accuracy when the level of environmental support is high. Although not detected within the ANOVA analysis, this does nonetheless raise the possibility that in certain situations older adults may show an advantage over young adults in predicting their future recognition of unrecalled items.

6.2.3 Recollection and FOK

A further consideration when examining memory and FOK is the potential relationship between recollection and FOK. Retrieval from memory is thought to involve two processes; recollection and familiarity (Yonelinas, 2002). Recollection is associated with the retrieval of specific contextual details in addition to the target memory, and thus relates to episodic memory, while familiarity is a general feeling that the item has been encountered before, and is associated with semantic memory (Tulving, 1972, 1985b). Souchay et al. (2007) proposed that recollection and feeling of knowing may be based on the same memory details. For example, remembering the colour a word was presented in would provide a contextual detail that increases the richness of successful memory retrieval, but if retrieval failed it would also provide relevant partial information relating to the target that could increase the FOK experience. In contrast with previous findings (Hicks & Marsh, 2002; Souchay et al., 2007), no evidence was found for a relationship between FOK and recollection as measured with the Remember-Know procedure in this thesis.

The recollection literature has also been used to support the assumption of selective deficits in episodic FOKs in ageing. Recollection is typically impaired in ageing, whether measured by the R/K procedure (e.g. Bugaiska et al., 2007; Bunce, 2003; Perfect & Dasgupta, 1997) or by examining source monitoring (Dennis et al., 2008; Mitchell & Johnson, 2009; Thomas et al., 2011), whereas familiarity is intact. Combined with the previously observed relationship between FOK and recollection, age-associated impairments in recollection (related to episodic memory) with preserved familiarity (related to semantic memory) were used to infer a similar pattern of deficits within FOK in ageing (Souchay et al., 2007). Indeed, Souchay et al. (2007) observed this very pattern, with older adults giving reduced R responses and showing lower FOK accuracy compared to young adults. Experiments 2.1 and 3.2 examined R/K and FOK in ageing. Unlike previous findings, no age effect was observed on R/K responding within these studies; young and older adults reported similar levels of recollection. However age effects were observed on FOK accuracy, with a selective impairment of episodic FOKs in Experiment 2.1 and reduced magnitude of FOKs in Experiment 3.2. Thus in addition to the lack of correlations, ageing had contrasting effects on the two subjective states.

This discrepancy in findings could be due to the methodologies used. As previously noted with regards to the effects of learning on FOK, the type of experimental design used can influence the pattern of responding when looking at subjective experiences (Carroll & Nelson, 1993). Souchay et al. (2007) measured FOK and R/K responding with separate tasks, while the current thesis examined them within the same task. Although both lines of evidence ensured responses were obtained from the same participants, the experiments reported in this thesis obtained both measures within the same task as well. As discussed in Chapter 4, there is evidence to suggest that metacognitive ability may not be stable across tasks (Thompson & Mason, 1996). It may therefore be possible that measures of recollection also show variability across tasks. Thus a within subjects and within task comparison of these two subjective experiences, as presented here, may reflect more accurately the true nature of the relationship between FOK and R/K responding. Alternatively, asking participants to make an FOK judgement prior to making an R/K judgement may have altered their responding. Previous studies have suggested that differences in the instructions or procedures used when obtaining R/K responses can have a large impact on the pattern of results obtained (e.g. McCabe & Geraci, 2009), thus asking participants to judge their future recognition may have affected the reported recollective experience at recognition.

6.2.4 A signal detection theory model of FOK

The application of signal detection theory within the FOK literature not only raises questions about how best to measure FOK accuracy, but also causes us to consider the underlying nature of FOK itself. Traditionally, signal detection models of recognition memory have consisted of a target distribution and a distractor distribution, with a recognition criterion above which items are identified as being previously presented and below which items are identified as being newly encountered. To adapt this model to the FOK, first it needs to be considered whether the FOK can be conceptualised as a single distribution ranging from low to high, or whether FOK is composed of two distributions, a positive FOK and a negative FOK. A small number of studies have shown support for this dual-process approach to FOK, with evidence from both experimental (Glucksberg & McCloskey, 1981; Klin et al., 1997) and neuroimaging (Liu et al., 2007; Luo, 2003) work.

However the predominant approach to FOK has assumed it lies on a single distribution ranging from low to high, with a decision criterion above which a positive FOK is given and below which a negative FOK is given. Therefore this conceptualisation of FOK was used in the model presented.

The precise construction of the model used reflects the predominant theories as to how FOKs are formed. The metacognitive account proposes that the meta-level consists not only of knowledge about how factors can influence memory, but also contains a representation of what is occurring at the object level (Nelson & Narens, 1990, 1994). Thus when an FOK is made, it is based on the representation of the target item in the meta-level. However, this representation is not perfect; the meta-level does not have direct access to the object level but instead must infer the contents of the object level. This is achieved by access to partial information relevant to the target which is activated during the failed recall attempt (Koriat & Levy-Sadot, 2001; Koriat, 1993, 1997). If the partial information retrieved is not diagnostic of the true strength of the target memory, then the representation formed of the target item in the meta-level will lead to inaccurate FOK predictions. Error in the FOK judgement may therefore be considered as to the level of discrepancy between the contents of the object level and the representation of the object level held in the meta-level.

The model proposed therefore consisted of a target item distribution, relating to the memory strength of the unrecalled items, and an FOK distribution, relating to the meta-level representation of the target item distribution. The amount of rotation between the FOK and target items distributions represented the error resulting from reliance on inferential processes. Greater rotation leads to a greater disparity between the two distributions, resulting in less accurate FOK predictions. In addition to the two distributions, the model also contained a decision criterion for each distribution, above which positive decision would be made and below which a negative decision would be made. Thus not only can the model provide a measure of accuracy, but it can also provide a measure of the bias in FOK responding; how likely it is a participant gives a Yes or a No FOK judgement.

The model was tested with both simulated data and experimental data from Experiments 2.1 and 2.2. The measure of rotation was found to be closely related to the gamma correlation, suggesting that amount of rotation measures a similar concept of accuracy to the widely accepted current measure of FOK accuracy. The model also provides an additional measure, bias, which is not typically considered in FOK research. A number of limitations exist with the model, such as the effect of missing data and the high computational expense. Nonetheless the model provides a convincing portrayal of the FOK which fits with current theoretical understanding of how FOKs are formed. In addition, the model makes explicit the underlying assumption throughout the FOK literature that the FOK consists of a single process. An interesting avenue of future research would be to consider the viability of a two process model of FOK as suggested by some researchers, and which of these models provides the closest account of the existing experimental data.

6.2.5 Judgements of retrieval failure and global FOKs

As well as applying existing measures to new areas of research, the development of novel measures can also drive metacognitive theory. Two new global measures and their relationship with memory manipulations were examined; the judgement of retrieval failure (JORF) and global FOK (GFOK). The JORF is a retrospective judgement of the amount of information which has failed to be retrieved, while the GFOK is a prediction of how much of this information will be recognised at a later date. JORF and GFOK were also found to be preserved in early stage dementia.

The implementation of a 14 week delay in Experiment 5.1 had clear effects on memory accuracy, with retrieval failures increasing both at recall and recognition test. These effects were mirrored in the metacognitive judgements obtained. JORF predictions increased as retrieval failure at recall increased, and GFOK predictions decreased as fewer items were correctly recognised after the delay. Older adults and patients diagnosed with early stage dementia were also found to give JORFs and GFOKs which were sensitive to the level of retrieval failure. Importantly, although the patients with dementia showed memory impairments compared to older adults, as expected, they also gave higher JORF and lower GFOK ratings, showing that they were aware of this memory impairment.

The category paradigm used provides a potential explanation of how these judgements are made. For the JORF, participants may have employed a generate-recognise strategy to aid their recall performance (Anderson & Bower, 1972). By generating exemplars, participants can then judge whether the generated item seems familiar due to being recently encountered or simply due to being a known exemplar. Those items which they are unsure of will then contribute to the JORF assessment. As the delay increases, the ability to discriminate exemplars from those presented to those that constitute the semantic category will become more difficult (Brainerd & Reyna, 2002; Gernsbacher, 1985). This leads to source memory errors (Johnson et al., 1993), where the familiarity of a generated item is attributed to it being recently encountered rather than to it merely being part of the category. Thus the discrepancy between JORF and retrieval failure, and the potential interaction between JORF and delay, may be due to these source memory errors.

The GFOK may also work on a similar basis. Although the GFOK may not necessarily rely on access to partial information in the same way the item-by-item FOK does, it would be reasonable to assume a similar process underlies the GFOK prediction. The activation of the semantic category and the exemplars generated when making the JORF may therefore also contribute to the GFOK. Again error will arise if this partial information is attributed to recently encountered items rather than general examples from the category. The contribution of the generate-recognise strategy and source memory errors would be easily verified experimentally by asking participants to report on their strategy use.

The JORF was also considered for semantic and episodic memory in Experiment 5.2. While the methodology employed precludes any firm conclusions, there may be a difference in resolution of semantic JORFs and episodic JORFs, as has been observed for semantic and episodic FOKs. Further consideration of this is required with a more highly controlled task design to remove potential confounds of set size. The difficulty with measuring a semantic JORF is the level of expertise the participant may have in the semantic category chosen. For example, while the participants used in the study had limited knowledge of the periodic table and thus gave a range of JORFs, if participants were chemistry graduates they may know the precise number of elements independently of the elements themselves, and so give perfect JORFs irrespective of their ability to recall or recognise each element. This represents the difference between known unknowns and unknown unknowns discussed by Hampton, Aina, Andersson, Mirza, & Parmar (2012). So while semantic JORF was less accurate than episodic JORF as measured in Experiment 5.2, it is possible for the reverse to also be observed.

6.3 Theoretical implications and methodological considerations

The findings of this thesis emphasise the complexity of the relationship between memory and feeling of knowing experiences. Manipulations of memory accuracy do not necessarily affect FOK accuracy. However, FOK judgements are sensitive to these manipulations, as evidenced by adjustments in magnitude in line with increases and decreases in memory performance.

The memory constraint hypothesis (MCH, MacLaverty & Hertzog, 2009) proposed that observations of an age-related deficit in episodic FOK accuracy in ageing can be attributed to deficits in encoding. Although it does not propose that FOKs generally have direct access to the memory trace, it does propose that memory factors are responsible for the age associated episodic FOK deficit rather than a deficit in metacognitive processes themselves. It assumes that older adults are unable to encode the information to the same level as young adults, leading to lower memory performance and less diagnostic partial information, subsequently affecting FOK accuracy. The findings of this thesis do not support this assumption. The effect of ageing was considered over a variety of studies with a variety of tasks, none of which observed deficits in FOK accuracy in conjunction with deficits in memory performance. Rather, older adults were found on many occasions to have equivalent or higher memory accuracy, but still show evidence of a selective episodic FOK deficit. The question then remains as to what is the driving force behind this deficit, whether it can be fully attributed to deficits in associated processes such as executive function, or also reflects problematic metacognitive monitoring.

An issue related to the MCH and which has not been fully explored in this thesis is the idea that recall and recognition rely on slightly different processes. This is especially of interest in ageing, where age-associated deficits are observed for recall while recognition remains relatively preserved (Botwinick & Storandt, 1980; Craik & McDowd, 1987). If recall processes are impaired, the partial information accessed during a failed retrieval attempt may also be impaired, leading to a reduction in the FOK predictions. However, if recognition is preserved, and the participant is not aware of this preserved accuracy, the FOK prediction given will not reflect performance at the recognition task. Based on reduced recall and access to partial information at recall the participant may assume memory more generally is impaired, rather than being specific to the current test of memory, leading to inaccuracies in the FOK judgement. Thus not only can the type of recognition task itself influence the FOK (Leonesio & Nelson, 1990; Schwartz & Metcalfe, 1994), but the differences between recall and recognition could also have a major impact. Further investigation of the relative impacts of ageing and experimental manipulations on recall and recognition separately and their relation to FOK may provide new insights in to the relationship between memory and metacognitive judgements.

Koriat, Bjork, Sheffer, & Bar (2004) proposed a distinction between theory-based and experience-based heuristics for metacognitive judgements. This distinction was established using the judgement of learning methodology, as it allowed for easier dissociation between the two processes. FOKs in contrast will always contain some level of experience-based information, as they occur following an unsuccessful recall attempt. However, in the present thesis it was demonstrated that the weighting given to each type of heuristic could be manipulated through the implementation of a delay, subsequently affecting FOK magnitude. This therefore confirmed the applicability of this distinction to metacognitive judgements beyond the JOL, and further identifies the qualitative nature of the partial information on which the FOK is based.

Perhaps the most striking theoretical implication of the present thesis is the signal detection based model of FOK outlined in Chapter 4. The model provides an integrated conceptualisation of a number of existing assumptions in the FOK literature and also provides a measure of accuracy which fits with the currently adopted measurement of the FOK. For example, the majority of researchers assume the FOK varies on a single dimension from low to high, with a threshold at which a positive FOK judgement is made and below which a negative FOK judgement is made. This assumption of a single distribution has allowed for the adoption of rating scales for FOK judgements, enabling finer measurement of participants' ability to discriminate between items (Goldsmith et al., 2002; Koriat, 1975). The model presented formalises this conception, with the FOK as a single distribution and a decision criterion to determine positive and negative FOK responding. In addition, one of the central tenets of the metacognitive approach is the assumption that the meta level holds a representation of the object level, against which the decision is made as to whether the desired learning goal has been sufficiently met (Nelson & Narens, 1990, 1994). The model also works on this assumption, that the FOK distribution is a representation of the unrecalled item distribution. The level of agreement between the two distributions can be measured in terms of the angle of rotation, which in turn provides a measure of accuracy similar to the established measure of accuracy within the literature. The model can also account for the effect of partial information contributing to the FOK judgement (Koriat & Levy-Sadot, 2001; Koriat, 1993). As the quantity and quality of partial information increases, the representation built will reflect the item distribution, leading to lower rotation and therefore increased accuracy. The model thus fits with the existing literature as well as providing an extension to the literature by introducing signal detection principles to the FOK.

One limitation of the model, and a limitation with FOK research generally, is the problem of missing data. If participants do not give a response for every possible outcome, it is not possible to ascertain their ability to discriminate between items which will be recognised in future and items which will not (Nelson, 1984). Although Souchay, Isingrini, & Gil (2002) proposed the adoption of an adjustment to give an approximation of predictive accuracy, this requires certain assumptions regarding the nature of the FOK judgement which may or may not fit with the conception of FOK the particular researcher has. While this can be controlled for with a semantic FOK task by continuing testing until all response options are used, this method is not available for an episodic FOK task. The problem cannot be circumnavigated by giving participants more items to learn in the hope that the next test session will provide the missing response outcomes, as experience with the task may lead participants to improve in accuracy between the two test sessions even though different items are used at encoding. With this in mind, the adoption of an adjustment would appear to be the most suitable course of action, with the caveat that the measure of accuracy adopted is an approximation of true accuracy.

A related issue with episodic FOK research is the limited number of responses on which a measure of FOK accuracy is established (Nelson, 1984). Participants can only encode a limited amount of information within a test session. If too few items are presented, recall will be too great and thus very few items will form the basis of the FOK. Likewise, if too many are presented, recall will be too low, and this may itself lead to a reduction in FOK magnitude, less willingness in participants to predict their performance, or even to attempt to recall items. The main approach to reduce these issues is to pilot materials prior to testing to establish their level of difficulty. However this approach is not failsafe, and despite piloting materials, floor and ceiling effects can still occur. Thus the researcher should consider these limitations when interpreting their results, as demonstrated within this thesis.

6.4 Future directions

A recent development in the metacognitive literature is the idea that metacognition is composed of two levels. One is relatively automatic and based on intuitions and experiences, whereas the other is higher order and meta-representational. In this thesis, this idea was not explicitly tested, but nonetheless the data can speak a little as to whether to not this theory explains the formation of an FOK. Some global judgements such as the GFOK and JORF for instance are clearly dependent more on the meta-representational level. To judge, as a number, the amount of data missing from a list, one presumably needs a mental representation of that list. Other judgements, for instance the classic FOK before and after a delay, presumably better captures the idea that feelings and familiarity can be used to make an FOK. In fact, it might be assumed that metacognitive aspects such as beliefs and expectations remain stable over the two hour delay, but what changes is the access to and experience of memory contents.

The signal detection model of FOK also relates to this delineation of automatic and higher order representation in metacognition. The model provides a possible conceptualisation of how this higher order representation may function, and also fits with current theoretical approaches as to how experiences and beliefs can influence metacognitive accuracy. Thus the model characterises how the intuitions and experiences contribute to the higher order representation of the contents of memory. Furthermore, the model raises another avenue of theoretical and experimental consideration: whether the FOK is considered as single process or dual process. The implicit assumption within metacognition is that sensations range from low to high, in that a low level sensation relates to a negative FOK judgement while a strong sensation relates to a positive FOK. The model proposed follows this assumption. A number of findings also indicate that the FOK may be composed of two separate processes, one which characterises negative FOK and one which characterises positive FOKs (Cheng, 2010; Glucksberg & McCloskey, 1981; Liu et al., 2007; Luo, 2003). The signal detection framework adopted could allow for these two opposing views to be explored, as has previously occurred within the recognition memory literature.

The impact of ageing in metacognition remains an important tool for further understanding metacognitive predictions. The findings of this thesis point to a selective, albeit mild, deficit in episodic FOK accuracy in ageing. Although not directly compared with young adults, the JORF and GFOK measures in older adults appear to also be preserved, with judgements similar to actual performance. As the FOK, JORF and GFOK all take place at retrieval failure, it is intriguing to note that while one is impaired, the others may be preserved. Whether this is due to methodological differences or due to differences in the underlying processes could not only shed light on the judgements themselves, but also inform the relationship between memory and metacognitive experience. The selective deficits in ageing may also speak to the theoretical assumption outlined above that metacognition may consist of two levels. It is possible that one level may be impaired in ageing while the second level may not. Tasks which rely primarily on the affected level may therefore show deficits in ageing, while task which rely more heavily on the preserved level may be preserved.

One final point of note is the possibility of training metacognitive abilities. Both young and older adults were able to improve the accuracy of their FOK predictions over repeated learning trials. Thus as experience with the task increased, so too did their ability to identify which unrecalled items would be subsequently recognised and which would not. It would be interesting to explore whether this improvement in accuracy would transfer across tasks. From a clinical perspective this is a promising finding, especially as older adults were also found to improve in accuracy despite exhibiting a slight FOK deficit. If metacognitive abilities can be trained, then patient groups who exhibit deficits within this domain may also be able to benefit from training to improve their memory regulation.

6.5 Conclusions

This thesis aimed to shed light on the relationship between memory accuracy and sensations at retrieval failure. The studies presented here have expanded on the existing literature with the feeling of knowing paradigm and suggested that while FOK is sensitive to manipulations of memory, these do not necessarily impact the accuracy of the predictions made. Findings also suggest that episodic FOK is selectively impaired in ageing while semantic FOKs are preserved, and that both young and older adults have the capacity to improve their episodic FOK accuracy.

Evidence for preserved memory awareness in early stage dementia was also found for two novel measures of metacognitive sensations. Finally, the thesis presented a novel model of the relationship between the feeling of knowing and memory, providing a new avenue of conceptualising and measuring feeling of knowing accuracy.

References

Allen-Burge, R., & Storandt, M. (2000). Age equivalence in feeling-of-knowing experiences. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *55*(4), 214–223.

Altarriba, J., & Mathis, K. M. (1997). Conceptual and Lexical Development in Second Language Acquisition. *Journal of Memory and Language*, *36*(4), 550–568.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97–123.

Anderson, J. W., & Schmitter-Edgecombe, M. (2010). Mild cognitive impairment and feeling-of-knowing in episodic memory. *Journal of Clinical and Experimental Neuropsychology*, *32*(5), 505–514.

Anderson, N. D., & Craik, F. I. M. (2000). Memory in the ageing brain. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory*. New York: Oxford University Press.

Ansell, E. L., & Bucks, R. S. (2006). Mnemonic anosognosia in Alzheimer's disease: A test of Agnew and Morris (1998). *Neuropsychologia*, *44*(7), 1095–102.

Babinski, M. J. (1914). Contribution à l'etude des troubles mentaux dans l'hémiplégie organique cerebrale (Anosognosia). *Revue Neurologique*, *22*, 845–848.

Bäckman, L., & Karlsson, T. (1985). The relation between level of general knowledge and feeling-of-knowing: An adult age study. *Scandinavian Journal of Psychology*, *26*(3), 249–258.

Bäckman, L., & Lipinska, B. (1993). Monitoring of general knowledge: Evidence for preservation in early Alzheimer's disease. *Neuropsychologia*, *31*(4), 335–345.

Bacon, E., Danion, J. M., Kauffmann-Muller, F., Schelstraete, M. A., Bruant, A., Sellal, F., & Grange, D. (1998). Confidence level and feeling of knowing for episodic and semantic memory: An investigation of lorazepam effects on metamemory. *Psychopharmacology*, *138*(3-4), 318–325.

Bacon, E., Danion, J.-M., Kauffmann-Muller, F., & Bruant, A. (2001). Consciousness in schizophrenia: A metacognitive approach to semantic memory. *Consciousness and Cognition*, *10*(4), 473–484.

Bacon, E., & Izaute, M. (2009). Metacognition in schizophrenia: Processes underlying patients' reflections on their own episodic memory. *Biological Psychiatry*, *66*(11), 1031–1037.

Bacon, E., Schwartz, B. L., Paire-Ficout, L., & Izaute, M. (2007). Dissociation between the cognitive process and the phenomenological experience of TOT: Effect of the anxiolytic drug lorazepam on TOT states. *Consciousness and Cognition*, *16*(2), 360–73.

Bahrick, H. P. (2008). Thomas O. Nelson: His life and comments on the implications of his functional view on metacognitive memory monitoring. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 1–7). New York: Psychology Press.

Baran, B., Tekcan, A. I., Gürvit, H., & Boduroglu, A. (2009). Episodic memory and metamemory in Parkinson's disease patients. *Neuropsychology*, *23*(6), 736–745.

Barrett, A. M., Eslinger, P. J., Ballentine, N. H., & Heilman, K. M. (2005). Unawareness of cognitive deficit (cognitive anosognosia) in probable AD and control subjects. *Neurology*, *64*(4), 693–699.

Benjamin, A. S., & Craik, F. I. M. (2001). Parallel effects of aging and time pressure on memory for source: Evidence from the spacing effect. *Memory & Cognition*, *29*(5), 691–697.

Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 73–93). New York: Psychology Press.

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review*, *116*(1), 84–115.

Berry, J. M., West, R. L., & Dennehey, D. M. (1989). Reliability and validity of the Memory Self-Efficacy Questionnaire. *Developmental Psychology*, *25*(5), 701–713.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.

Blake, M. (1973). Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *12*(3), 311–319.

Botwinick, J., & Storandt, M. (1980). Recall and recognition of old information in relation to age and sex. *Journal of Gerontology*, *35*(1), 70–6.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-Trace Theory and False Memory. *Current Directions in Psychological Science*, *11*(5), 164–169.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *The British Journal of Clinical Psychology*, *21*(1), 1–16.

Brown, A. L. (1978). Knowing when, where and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in Instructional Psychology*. Hillsdale, NJ: Erlbaum.

Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*, 325–337.

Buffet-Jerrott, S. E., & Stewart, S. H. (2002). Cognitive and sedative effects of benzodiazepine use. *Current Pharmaceutical Design*, *8*, 45–58.

Bugaiska, A., Clarys, D., Jarry, C., Taconnat, L., Tapia, G., Vanneste, S., & Isingrini, M. (2007). The effect of aging in recollective experience: The processing speed and executive functioning hypothesis. *Consciousness and Cognition*, *16*(4), 797–808.

Bunce, D. (2003). Cognitive support at encoding attenuates age differences in recollective experience among adults of lower frontal lobe function. *Neuropsychology*, *17*(3), 353–361.

Bunnell, J. K., Baken, D. M., & Richards-Ward, L. A. (1999). The effect of age on metamemory for working memory. *New Zealand Journal of Psychology*, *28*(1), 23–29.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*(5), 542–579.

Burke, D., Worthley, J. S., & Wade, E. (1988). I’ll never forget what's-her-name: Aging and tip of the tongue experiences in everyday life. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical Aspects of Memory: Current Research and Issues* (pp. 113–118). New York: Wiley.

Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, *24*(5), 654–663.

Caldwell, J. I., & Masson, M. E. J. (2001). Conscious and unconscious influences of memory for object location. *Memory & Cognition*, *29*(2), 285–295.

Carroll, M., & Nelson, T. O. (1993). Effect of overlearning on the feeling of knowing is more detectable in within-subject than in between-subject designs. *American Journal of Psychology*, *106*(2), 227–235.

Carroll, M., & Simington, A. (1986). The effects of degree of learning, meaning, and individual differences on the feeling-of-knowing. *Acta Psychologica*, *61*, 3–16.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language, 49*(2), 231-248.

Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory : A critical examination. *Child Development*, *53*(1), 11–28.

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, *24*(4), 403–416.

Cheng, C.-M. (2010). Accuracy and stability of metacognitive monitoring: A new measure. *Behavior Research Methods*, *42*(3), 715–732.

Clare, L., Wilson, B. A., Roth, I., & Hodges, J. R. (2002). Assessing awareness in early-stage Alzheimer's disease: Development and piloting of the Memory Awareness Rating Scale. *Neuropsychological Rehabilitation*, *12*(4), 341–362.

Clare, L., & Woods, R. T. (2004a). Cognitive training and cognitive rehabilitation for people with early-stage Alzheimer's disease: A review. *Neuropsychological Rehabilitation*, *14*(4), 385–401.

Clare, L., & Woods, R. T. (2004b). Cognitive training and cognitive rehabilitation for people with early-stage Alzheimer's disease: A review. *Neuropsychological Rehabilitation*, *14*(4), 385–401.

Clark, S. E., & Godfrey, R. D. (2009). Eyewitness identification evidence and innocence risk. *Psychonomic Bulletin & Review*, *16*(1), 22–42.

Clarys, D., Bugaiska, A., Tapia, G., & Baudouin, A. (2009). Ageing, remembering, and executive function. *Memory*, *17*(2), 158–168.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

Connor, L. T., Balota, D. A., & Neely, J. H. (1992). On the relationship between feeling of knowing and lexical decision: Persistent subthreshold activation or topic familiarity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 544–554.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, *12*(1), 50–71.

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General*, *126*(4), 393–413.

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2006). Source memory in the absence of successful cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 828–835.

Correa, D. D., Graves, R. E., & Costa, L. (1996). Awareness of memory deficit in Alzheimer's Disease patients and memory-impaired older adults. *Aging, Neuropsychology, and Cognition*, *3*(3), 215–228.

Costermans, J., Lories, G., & Ansay, C. (1992). Confidence level and feeling of knwoing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 142–150.

Craik, F. I. M., Anderson, N. D., Kerr, S. A., & Li, K. Z. H. (1995). Memory changes in normal ageing. In A. D. Baddeley, B. A. Wilson, & F. N. Watts (Eds.), *Handbook of Memory Disorders* (pp. 211–241). Chichester: Wiley.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 474–479.

Dalla Barba, G., Parlato, V., Iavarone, A., & Boller, F. (1995). Anosognosia, intrusions and "frontal" functions in Alzheimer's Disease and depression. *Neuropsychologia*, *33*(2), 247–259.

Dennis, N. A., Hayes, S. M., Prince, S. E., Madden, D. J., Huettel, S. A., & Cabeza, R. (2008). Effects of aging on the neural correlates of successful item and source memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 791–808.

Dixon, R. A, & Hultsch, D. F. (1983). Structure and development of metamemory in adulthood. *Journal of Gerontology*, *38*(6), 682–8.

Dobbins, I. G., & Han, S. (2007). What constitues a model of item-based memory decision-making? In A. Benjamin & B. Ross (Eds.), *The Psychology of Learning and Motivation: Volume 48*. London: Elsevier.

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*(1), 155–61.

Dodson, C. S., & Schacter, D. L. (2002). Aging and strategic retrieval processes: Reducing false memories with a distinctiveness heuristic. *Psychology and Aging*, *17*(3), 405–415.

Duke, L. M., Seltzer, B., Seltzer, J. E., & Vasterling, J. J. (2002). Cognitive components of deficit awareness in Alzheimer's disease. *Neuropsychology*, *16*(3), 359–369.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition*, *25*(5), 691–700.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (pp. 249–276). Mahwah, NJ: Erlbaum.

Eagle, M. N. (1967). The effect of learning strategies upon free recall. *American Journal of Psychology*, *80*(3), 421–425.

Ebbinghaus, H. (1913). Retention as a function of the order of succession of the members of the series. In H. A. Ruger & C. E. Bussenius (Eds.), *Memory: A contribution to experimental psychology*. New York: Teacher's College Press.

Engle, R. W., & Bukstel, L. (1978). Memory processes among bridge players of differing expertise. *American Journal of Psychology*, *91*(4), 673–689.

Eysenck, M. W. (1979). The feeling of knowing a word's meaning. *British Journal of Psychology*, *70*, 243–251.

Ezrachi, O., Ben-Yishay, Y., Kay, T., Diller, L., & Rattok, J. (1991). Predicting employment in trauimatic brain injury following neuropsychological rehabilitation. *Journal of Head Trauma and Rehabilitation*, *6*(3), 71–84.

Feehan, M., Knight, R. G., & Partridge, F. M. (1991). Cognitive complaint and test performance in elderly patients suffering depression or dementia. *International Journal of Geriatric Psychiatry*, *6*, 287–293.

Feher, E. P., Mahurin, R. K., Inbody, S. B., Crook, T. H., & Pirozzolo, F. J. (1991). Anosognosia in Alzheimer's Disease. *Nueorpsychiatry, Neuropsychology and Behavioral Neurology*, *4*(2), 136–146.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and Cognition*, *9*(2 Pt 1), 288–307.

Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition*, *36*(4), 813–821.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, *34*(10), 906–911.

Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, *1*, 324–340.

Fleming, J. M., Strong, J., & Ashton, R. (1996). Self-awareness of deficits in adults with traumatic brain injury: How best to measure? *Brain Injury*, *10*(1), 1–15.

Fleming, J. M., Strong, J. S., & Ashton, R. A. (1998). Cluster analysis of self-awareness levels in adults with traumatic brain injury and relationship to outcome. *Journal of Head Trauma and Rehabilitation*, *13*(5), 39–51.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Friedman, D., de Chastelaine, M., Nessler, D., & Malcolm, B. (2010). Changes in familiarity and recollection across the lifespan: An ERP perspective. *Brain Research*, *1310*, 124–141.

Froger, C., Bouazzaoui, B., Isingrini, M., & Taconnat, L. (2012). Study time allocation deficit of older adults: The role of environmental support at encoding? *Psychology and Aging*, *27*(3), 577–588.

Gallo, D. A, Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic bulletin & review*, *8*(3), 579–86.

Gardiner, J M. (1988). Recognition failures and free-recall failures: Implications for the relation between recall and recognition. *Memory & cognition*, *16*(5), 446–51.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, *I*(3), 213–216.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory*, *10*(2), 83–98.

Gernsbacher, M. A. (1985). Surface information Loss in Comprehension. *Cognitive Psychology*, *17*, 324–363.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506–528.

Gilewski, M. J., Zelinski, E. M., & Schaie, K. W. (1990). The Memory Functioning Questionnaire for assessment of memory complaints in adulthood and old age. *Psychology and Aging*, *5*(4), 482–490.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*(6), 597–602.

Glucksberg, S., & McCloskey, M. (1981). Decisions about ignorance: Knowing that you don’t know. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(5), 311–325.

Goh, J. O., Beason-held, L. L., An, Y., Kraut, M. A., & Resnick, S. M. (2013). Frontal function and executive processing in older adults: Process and region specific age-related longitudinal functional changes. *NeuroImage*, *69*, 43–50.

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General*, *131*(1), 73–95.

Gonzalez, R., & Nelson, T. O. (1996). Measuring ordinal association in situations that contain tied scores. *Psychological Bulletin*, *119*(1), 159–165.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*(268), 732–764.

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

Green, J., Goldstein, F. C., Sirockman, B. E., & Green, R. C. (1993). Variable awareness of deficits in Alzheimer's Disease. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *6*(3), 159–165.

Griffin, D., & Buehler, R. (1999). Frequency, probability, and prediction: Easy solutions to cognitive illusions? *Cognitive Psychology*, *38*(1), 48–78.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, *24*, 411–435.

Gruneberg, M. M. (1983). Strategies in human memory. In A. R. Mayes (Ed.), *Memory in Animals and Man*. London: Van Nostrand.

Gruneberg, M. M., & Monks, J. (1974). Feeling of knowing and cued recall. *Acta Psychologica*, *38*, 257–265.

Gruneberg, M. M., Monks, J., & Sykes, N. (1977). Some methodological problems with feeling of knowing studies. *Acta Psychologica*, *41*, 365–371.

Gruneberg, M. M., & Sykes, R. N. (1978). Knowledge and retention: The feeling of knowing and reminiscence. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical Aspects of Memory* (pp. 189–196). New York: Academic Press.

Guttentag, R. E., & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory and Language*, *37*(4), 502–516.

Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 691–702.

Halamish, V., McGillivray, S., & Castel, A. D. (2011). Monitoring one's own forgetting in younger and older adults. *Psychology and Aging*, *26*(3), 631– 635.

Hampton, J. A, Aina, B., Andersson, J. M., Mirza, H. Z., & Parmar, S. (2012). The Rumsfeld effect: The unknown unknown. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 340–55.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208–216.

Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, *6*, 685–691.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values ofd'. *Behavior Research Methods, Instruments, and Computers*, *27*(1), 46–51.

Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1210–30.

Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General*, *127*(3), 251–268.

Herrman, D. J., & Neisser, U. (1978). An inventory of everyday memory. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical Aspects of Memory* (pp. 35–51). New York: Academic Press.

Herrmann, D. J. (1982). Know thy memory: The use of questionnaires to assess and study memory. *Psychological Bulletin*, *92*(2), 434–452.

Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & Cognition*, *38*(6), 771–784.

Hertzog, C., Saylor, L. L., Fleece, A. M., & Dixon, R. A. (1994). Metamemory and aging: Relations between predicted , actual and perceived memory task performance. *Aging, Neuropsychology, and Cognition*, *1*(3), 203–237.

Hicks, J. L., & Marsh, R. L. (2002). On predicting the future states of awareness for recognition of unrecallable items. *Memory & Cognition*, *30*(1), 60–66.

Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 503–508.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1321–1330.

Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition*, *27*(4), 657–664.

Hosey, L. A., Peynircioğlu, Z. F., & Rabinovitz, B. E. (2009). Feeling of knowing for names in response to faces. *Acta Psychologica*, *130*(3), 214–224.

Howard, C. E., Andrés, P., Broks, P., Noad, R., Sadler, M., Coker, D., & Mazzoni, G. (2010). Memory, metamemory and their dissociation in temporal lobe epilepsy. *Neuropsychologia*, *48*(4), 921–932.

Ivory, S., Knight, R. G., Longmore, B. E., & Caradoc-Davies, T. (1999). Verbal memory in non-demented patients withidiopathic Parkinsons disease. *Neuropsychologia*, *37*(7), 817–828.

Izaute, M., & Bacon, E. (2006). Effects of the amnesic drug lorazepam on complete and partial information retrieval and monitoring accuracy. *Psychopharmacology*, *188*(4), 472–81.

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 21–38.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology*, *17*(1), 3–11.

Jennings, J., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging*, *8*(2), 283–293.

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

Jones, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition*, *17*(1), 69–76.

Jurado, M. A., Junqué, C., Vallejo, J., Salgado, P., & Grafman, J. (2002). Obsessive-compulsive disorder (OCD) patients are impaired in remembering temporal order and in judging their own performance. *Journal of Clinical and Experimental Neuropsychology*, *24*(3), 261–269.

Kelemen, W. L. (2000). Metamemory cues and monitoring accuracy: Judging what you know and what you will know. *Journal of Educational Psychology*, *92*(4), 800–810.

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, *28*(1), 92–107.

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language*, *35*, 157–175.

Keren, G., & Teigen, K. H. (2001). Why is p=.90 better than p = .70? Preference for definitive predictions by lay consumers of probability judgements. *Psychonomic Bulletin & Review*, *8*(2), 191–202.

Kikyo, H., & Miyashita, Y. (2004). Temporal lobe activations of “feeling-of-knowing” induced by face-name associations. *NeuroImage*, *23*(4), 1348–1357.

Kikyo, H., Ohki, K., & Miyashita, Y. (2002). Neural correlates for feeling-of-knowing: An fMRI parametric analysis. *Neuron*, *36*(1), 177–186.

Kim, J. (1971). Predictive measures of ordinal association. *American Journal of Sociology*, *76*(5), 891–907.

Klin, C. M., Guzmán, A. E., & Levine, W. H. (1997). Knowing that you don’t know: Metamemory and discourse processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1378–93.

Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 699–710.

Kok, A. (2000). Age-related changes in involuntary and voluntary attention as reflected in components of the event-related potential (ERP). *Biological Psychology*, *54*, 107–143.

Kolers, P. A., & Palef, S. R. (1976). Knowing not. *Memory & Cognition*, *4*(5), 553–8.

Koriat, A. (1975). Phonetic symbolism and feeling of knowing. *Memory & Cognition*, *3*(5), 545–548.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124*(3), 311–333.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*(4), 643–656.

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

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 34–53.

Koriat, A., Levy-Sadot, R., Edry, E., & de Marcas, S. (2003). What do we know about what we cannot remember? Accessing the semantic attributes of words that cannot be recalled. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1095–1105.

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*(4), 478–492.

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, *138*(4), 449–468.

Kozlowski, L. T. (1977). Effects of distorted auditory and of rhyming cues on retrieval of tip-of-the-tongue words by poets and nonpoets. *Memory & Cognition*, *5*(4), 477–481.

Krinsky, R., & Nelson, T. O. (1985). The feeling of knowing for different types of retrieval failure. *Acta Psychologica*, *58*, 141–158.

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. M. Gruneberg, P. Morris, & R. N. Sykes (Eds.), *Practical Aspects of Memory* (2nd ed.). London: Wiley.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections bewteen bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

Kroll, M. D., & Ford, M. L. (1992). The illusion of knowing, error detection, and motivational orientations. *Contemporary Educational Psychology*, *17*(4), 371–378.

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*(1), 23–55.

Lachman, J. L., Lachman, R., & Thronesbery, C. (1979). Metamemory through the adult life span. *Developmental Psychology*, *15*(5), 543–551.

Le Berre, A.-P., Pinon, K., Vabret, F., Pitel, A.-L., Allain, P., Eustache, F., & Beaunieux, H. (2010). Study of metamemory in patients with chronic alcoholism using a feeling-of-knowing episodic memory task. *Alcoholism, Clinical and Experimental Research*, *34*(11), 1888–1898.

Le Ny, J.-F., Denhiere, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarity in self-paced learning conditions. *Acta Psychologica*, *36*, 280–289.

Lemaire, P. (2010). Cognitive strategy variations during aging. *Current Directions in Psychological Science*, *19*(6), 363–369.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 464–467.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149–188.

Liberman, V. (2004). Local and global judgments of confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 729–732.

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology*, *42*, 333–376.

Lipinska, B., & Backman, L. (1996). Feeling of knowing in fact retrieval: Further evidence for preservation in early Alzheimer's Disease. *Journal of the International Neuropsychological Society*, *2*(4), 350–358.

Liu, Y., Su, Y., Xu, G., & Chan, R. C. K. (2007). Two dissociable aspects of feeling-of-knowing: Knowing that you know and knowing that you do not know. *Quarterly Journal of Experimental Psychology*, *60*(5), 672–680.

Luo, J. (2003). Neural correlates of "feeling-of-not-knowing": Evidence from functional MRI. *Chinese Science Bulletin*, *48*(2), 144.

Lupker, S. J., Harbluk, J. L., & Patrick, A. S. (1991). Memory for things forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 897–907.

Lyon, T. D., & Flavell, J. H. (1993). Young children's understanding of forgetting over time. *Child Development*, *64*(3), 789–800.

MacDougall, R. (1904). Recognition and recall. *The Journal of Philosophy, Psychology and Scientific Methods*, *1*(9), 229–233.

MacLaverty, S. N., & Hertzog, C. (2009). Do age-related differences in episodic feeling of knowing accuracy depend on the timing of the judgement? *Memory*, *17*(8), 860–873.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Mandler, G. (1967). Organisation and memory. In K. W. Spence & J. T. Spence (Eds.), *The Psychology of Learning and Motivation: Volume 1*. New York: Academic Press.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271.

Maril, A., Wagner, A. D., & Schacter, D. L. (2001). On the tip of the tongue: An event-related fMRI study of semantic retrieval failure and cognitive conflict. *Neuron*, *31*(4), 653–60.

Maril, A., Simons, J. S., Mitchell, J. P., Schwartz, B. L., & Schacter, D. L. (2003). Feeling-of-knowing in episodic memory: An event-related fMRI study. *NeuroImage*, *18*(4), 827–836.

Maril, A., Simons, J. S., Weaver, J. J., & Schacter, D. L. (2005). Graded recall success: An event-related fMRI comparison of tip of the tongue and feeling of knowing. *NeuroImage*, *24*(4), 1130–1138.

Marquié, J. C., & Huet, N. (2000). Age differences in feeling-of-knowing and confidence judgements as a function of knowledge domain. *Psychology and Aging*, *15*(3), 451–460.

Marquié, J. C., Jourdan-Boddaert, L., & Huet, N. (2002). Do older adults underestimate their actual computer knowledge? *Behaviour & Information Technology*, *21*(4), 273–280.

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 509–527.

Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1263–1274.

McCabe, D. P., & Geraci, L. D. (2009). The influence of instructions and terminology on the accuracy of remember-know judgments. *Consciousness and Cognition*, *18*(2), 401–413.

McCabe, D. P., & Soderstrom, N. C. (2011). Recollection-based prospective metamemory judgments are more accurate than those based on confidence: Judgments of remembering and knowing (JORKS). *Journal of Experimental Psychology: General*, *140*(4), 605–621.

McGillivray, S., & Castel, A. D. (2011). Betting on memory leads to metacognitive improvement by younger and older adults. *Psychology and Aging*, *26*(1), 137–142.

McGlynn, S. M., & Schacter, D. L. (1989). Unawareness of deficits in neuropsychological syndromes. *Journal of Clinical and Experimental Neuropsychology*, *11*(2), 143–205.

McGlynn, S. M., & Kaszniak, A. W. (1991). When metacognition fails: Impaired awareness of deficit in Alzheimer's Disease. *Journal of Cognitive Neuroscience*, *3*(2), 183–189.

Merritt, P., Hirshman, E., Hsu, J., & Berrigan, M. (2005). Metamemory without the memory: Are people aware of midazolam-induced amnesia? *Psychopharmacology*, *177*(3), 336–43.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 851–861.

Meyer, A. S., & Bock, K. (1992). The tip-of-the-tongue phenomenon: Blocking or partial activation? *Memory & Cognition*, *20*(6), 715–26.

Miner, A. C., & Reder, L. M. (1994). A new look at feeling of knowing: Its metacognitive role in regulating question answering. In J Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing About Knowing* (pp. 47–70). Cambridge, MA: MIT Press.

Ming, Z., Criswell, H. E., Yu, G., & Breese, G. R. (2006). Competing presynaptic and postsynaptic effects of ethanol on cerebellar purkinje neurons. *Alcoholism, Clinical and Experimental Research*, *30*(8), 1400–7.

Mintzer, M. Z. (2007). The acute effects of alcohol on memory: A review of laboratory studies in healthy adults. *International Journal on Disability and Human Development*, *6*(4), 397–403.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*(4), 638–677.

Morris, R. G., & Hannesdottir, K. (2004). Insight and cognition in Alzheimer's Disease. In R. G. Morris & J. Becker (Eds.), *The Cognitive Neuropsychology of Alzheimer's Disease*. Oxford: Oxford University Press.

Moulin, C. J. A. (2002). Sense and sensitivity: Metacognition in Alzheimer's Disease. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied Metacognition*. Cambridge, UK: Cambridge University Press.

Moulin, C. J. A., Perfect, T. J., & Jones, R. W. (2000a). Global predictions of memory in Alzheimer's Disease: Evidence for preserved metamemory monitoring. *Aging, Neuropsychology, and Cognition*, *7*(4), 230–244.

Moulin, C. J. A., Perfect, T. J., & Jones, R. W. (2000b). The effects of repetition on allocation of study time and judgements of learning in Alzheimer's disease. *Neuropsychologia*, *38*(6), 748–756.

Moulin, C. J. A., Perfect, T. J., & Jones, R. W. (2000c). Evidence for intact memory monitoring in Alzheimer's disease: Metamemory sensitivity at encoding. *Neuropsychologia*, *38*(9), 1242–1250.

Mueller, S. T., & Weidemann, C. T. (2008). Decision noise: An explanation for observed violations of signal detection theory. *Psychonomic Bulletin & Review*, *15*(3), 465–494.

Murphy, M. D., Sanders, R. E., Gabriesheski, A. S., & Schmitt, F. A. (1981). Metamemory in the aged. *Journal of Gerontology*, *36*(2), 185–193.

Murphy, M. D., Schmitt, F. a, Caruso, M. J., & Sanders, R. E. (1987). Metamemory in older adults: The role of monitoring in serial recall. *Psychology and Aging*, *2*(4), 331–339.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170–1187.

Naveh-Benjamin, M., Craik, F. I. M., & Ben-Shaul, L. (2002). Age-related differences in cued recall: Effects of support at encoding and retrieval. *Aging, Neuropsychology, and Cognition*, *9*(4), 276–287.

Nelson, T. O., McSpadden, M., Fromme, K., & Marlatt, G. a. (1986). Effects of alcohol intoxication on metamemory and on retrieval from long-term memory. *Journal of Experimental Psychology: General*, *115*(3), 247–54.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133.

Nelson, T. O. (1988). Predictive accuracy of the feeling of knowing across different criterion tasks and across different subject populations and individuals. In M

M Gruneberg (Ed.), *Practical Aspects of Memory: Current Research and Issues* (pp. 190–196). New York: Wiley.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*(2), 102–116.

Nelson, T. O., & Dunlosky, J. (1991). When people's Judgments of Learning (JOLs) are extremely accuract at predicting subsequent recall: The delayed JOL effect. *Psychological Science*, *2*(4), 267–271.

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory*, *2*(3), 325–335.

Nelson, T. O., Gerler, D., & Narens, L. (1984). Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General*, *113*(2), 282–300.

Nelson, T. O., Leonesio, R. J., Landwehr, R. S., & Narens, L. (1986). A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing versus the normative feeling of knowing versus base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 279–287.

Nelson, T. O., Leonesio, R. J., Shimamura, A. P., Landwehr, R. F., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(4), 279–288.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, *19*(3), 338–368.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 26, pp. 125–173). San Diego, CA: Academic Press.

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, *23*(1), 104–118.

Ownsworth, T., & Clare, L. (2006). The association between awareness deficits and rehabilitation outcome following acquired brain injury. *Clinical Psychology Review*, *26*(6), 783–95.

Ownsworth, T. L., McFarland, K., & Mc Young, R. (2000). Self-awareness and psychosocial functioning following acquired brain injury: An evaluation of a group support programme. *Neuropsychological Rehabilitation*, *10*(5), 465–484.

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology Review*, *15*(3), 105–30.

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30*(3), 331–347.

Pappas, B. A., Sunderland, T., Weingartner, H. M., Vitiello, B., Martinson, H., & Putnam, K. (1992). Alzheimer's Disease and feeling-of-knowing for knowledge and episodic memory. *Journal of Gerontology: Psychological Sciences*, *47*(3), 159–164.

Patterson, M. M., & Hertzog, C. (2010). The effects of age in four alternative forced-choice item and associative recognition tasks. *Psychology and Aging*, *25*(1), 235–238.

Perfect, T. J., & Dasgupta, Z. R. (1997). What underlies the deficit in reported recollective experience in old age? *Memory & Cognition*, *25*(6), 849–858.

Perfect, T. J., & Hollins, T. S. (1996). Predictive feeling of knowing judgements and postdictive confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology*, *10*, 371–382.

Perfect, T. J., & Hollins, T. S. (1999). Feeling of knowing judgments do not predict subsequent recognition performance for eyewitness memory. *Journal of Experimental Psychology: Applied*, *5*(3), 250–264.

Perfect, T. J., Mayes, A. R., Downes, J. J., & Eijk, R. Van. (1996). Does context discriminate recollection from familiarity in recognition memory? *The Quarterly Journal of Experimental Psychology*, *49*(3), 797–813.

Perfect, T. J., & Stollery, B. (1993). Memory and metamemory performance in older adults: One deficit or two ? *The Quarterly Journal of Experimental Psychology*, *46*(1), 119–135.

Perfect, T. J., Williams, R. B., & Anderton-Brown, C. (1995). Age differences in reported recollective experience are due to encoding effects, not response bias. *Memory*, *3*(2), 169–186.

Perrotin, A., Belleville, S., & Isingrini, M. (2007). Metamemory monitoring in mild cognitive impairment: Evidence of a less accurate episodic feeling-of-knowing. *Neuropsychologia*, *45*(12), 2811–2826.

Perrotin, A., Isingrini, M., Souchay, C., Clarys, D., & Taconnat, L. (2006). Episodic feeling-of-knowing accuracy and cued recall in the elderly: Evidence for double dissociation involving executive functioning and processing speed. *Acta Psychologica*, *122*(1), 58–73.

Perrotin, A., Tournelle, L., & Isingrini, M. (2008). Executive functioning and memory as potential mediators of the episodic feeling-of-knowing accuracy. *Brain and Cognition*, *67*(1), 76–87.

Peynircioğlu, Z. F., & Tekcan, A. İ. (2000). Feeling of knowing for translations of words. *Journal of Memory and Language*, *43*(1), 135–148.

Pinon, K., Allain, P., Kefi, M. Z., Dubas, F., & Le Gall, D. (2005). Monitoring processes and metamemory experience in patients with dysexecutive syndrome. *Brain and Cognition*, *57*(2), 185–188.

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, *86*, 227–272.

Price, J., Hertzog, C., & Dunlosky, J. (2010). Self-regulated learning in younger and older adults: Does aging affect metacognitive control? *Aging, Neuropsychology, and Cognition*, *17*(3), 329–359.

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, *17*(1), 39–57.

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*(3), 518–535.

Raz, N. (2000). Ageing of the brain and its impact on cognitive performance: Integration of structural and functional indings. In F. I. M. Craik & T. A. Salthouse (Eds.), *The Handbook of Ageing and Cognition* (2nd ed.). London: Lawrence Erlbaum Associates.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., … Acker, J. D. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex*, *15*(11), 1676–1689.

Read, J. D., & Bruce, D. (1982). Longitudinal tracking of difficult memory retrievals. *Cognitive Psychology*, *14*, 280–300.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, *19*(1), 90–138.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 435–451.

Reggev, N., Zuckerman, M., & Maril, A. (2011). Are all judgments created equal? An fMRI study of semantic and episodic metamemory predictions. *Neuropsychologia*, *49*(5), 1332–1342.

Roberts, L. S., & Rhodes, G. (1989). Knowing your limits: Expertise and feeling of knowing. *New Zealand Journal of Psychology*, *18*, 71–75.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814.

Ryan, M. P., Petty, C. R., & Wenzlaff, R. M. (1982). Motivated remembering efforts during tip-of-the-togue states. *Acta Psychologica*, *51*, 137–147.

Sacher, M., Taconnat, L., Souchay, C., & Isingrini, M. (2009). Divided attention at encoding: Effect on feeling-of-knowing. *Consciousness and Cognition*, *18*(3), 754–761.

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*, 35–54.

Schacter, D. L. (1983). Feeling of knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 39–54.

Schacter, D. L., & Worling, J. (1985). Attribute information and the feeling of knowing. *Canadian Journal of Psychology*, *39*(3), 467–475.

Schacter, D. L. (1983). Amnesia observed: Remembering and forgetting in a natural environment. *Journal of Abnormal Psychology*, *92*(2), 236–242.

Schacter, D. L, McLachlan, D. R., Moscovitch, M., & Tulving, E. (1986). Monitoring of recall performance by memory disorders patients. *Journal of Clinical and Experimental Neuropsychology*, *8*, 130.

Scheck, P., Meeter, M., & Nelson, T. O. (2004). Anchoring effects in the absolute accuracy of immediate versus delayed judgments of learning. *Journal of Memory and Language*, *51*(1), 71–79.

Schmitter-Edgecombe, M., & Anderson, J. W. (2007). Feeling of knowing in episodic memory following moderate to severe closed-head injury. *Neuropsychology*, *21*(2), 224–234.

Schmitter-Edgecombe, M., & Seelye, A. M. (2011). Predictions of verbal episodic memory in persons with Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *33*(2), 218–25.

Schnyer, D. M., Nicholls, L., & Verfaellie, M. (2005). The role of VMPC in metamemorial judgments of content retrievability. *Journal of Cognitive Neuroscience*, *17*(5), 832–846.

Schraw, G. (1995). Measures of feeling-of-knowing accuracy: A new look at an old problem. *Applied Cognitive Psychology*, *9*(4), 321–332.

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R., & Stroffolino, P. J. (1997). To calculate or not to calculate: A source activation confusion model of problem familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 3–29.

Schwartz, B. L. (2002). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical retrieval*. Mahwah, New Jersey: Erlbaum.

Schwartz, B. L. (2008). Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgments. *Memory & Cognition*, *36*(1), 9–19.

Schwartz, B. L., & Bacon, E. (2008). Metacognitive neuroscience. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 355–372). New York: Psychology Press.

Schwartz, B. L., & Frazier, L. D. (2005). Tip-of-the-tongue states and aging: Contrasting psycholinguistic and metacognitive perspectives. *The Journal of General Psychology*, *132*(4), 377–391.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling of knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1074–1083.

Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In A.P. Shimamura & J. Metcalf (Eds.), *Metacognition: Knowing About Knowing* (pp. 93–113). London, MIT Press.

Serra, M. J., & England, B. D. (2012). Magnitude and accuracy differences between judgements of remembering and forgetting. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2231–2257.

Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 452–460.

Sholl, A., Sankaranarayanan, A., & Kroll, J. F. (1995). Transfer betwen picture naming and tranlsation: A test of asymmetries in bilingual memory. *Psychological Science*, *6*(1), 45–49.

Skinner, E. I., & Fernandes, M. A. (2009). Age-related changes in the use of study context to increase recollection. *Aging, Neuropsychology, and Cognition*, *16*(4), 377–400.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*(5408), 1657–1661.

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.

Söderlund, H., Parker, E. S., Schwartz, B. L., & Tulving, E. (2005). Memory encoding and retrieval on the ascending and descending limbs of the blood alcohol concentration curve. *Psychopharmacology*, *182*(2), 305–17.

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review*, *27*(6), 799–811.

Souchay, C. (2007). Metamemory in Alzheimer's Disease. *Cortex*, *43*, 987–1003.

Souchay, C., & Isingrini, M. (2004). Age related differences in metacognitive control: Role of executive functioning. *Brain and Cognition*, *56*(1), 89–99.

Souchay, C., Isingrini, M., Clarys, D., Taconnat, L., & Eustache, F. (2004). Executive functioning and judgment-of-learning versus feeling-of-knowing in older adults. *Experimental Aging Research*, *30*(1), 47–62.

Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology*, *14*(2), 299–309.

Souchay, C., Isingrini, M., & Gil, R. (2002). Alzheimer's disease and feeling-of-knowing in episodic memory. *Neuropsychologia*, *40*(13), 2386–2396.

Souchay, C., & Moulin, C. J. A. (2009). Memory and consciousness in Alzheimer's Disease. *Current Alzheimer Research*, *6*(3), 186–195.

Souchay, C., Moulin, C. J. A., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness and Cognition*, *16*(4), 769–784.

Spellman, B. A., Bloomfield, A., & Bjork, R. A. (n.d.). Measuring memory and metamemory: Theoretical and statistical problems with assessing learning and using gamma to do so. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of Metamemory and Memory*. New York: Psychology Press.

Squire, L. R. (1984). Declarative and nondeclarative memory : Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, *4*(3), 232–243.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of learning and memory*, *82*(3), 171–7.

Sunderland, A., Watts, K., Baddeley, A. D., & Harris, J. E. (1986). Subjective memory assessment and test performance in elderly adults. *Journal of Gerontology*, *41*(3), 376–84.

Tauber, S. K., & Rhodes, M. G. (2010). Does the amount of material to be remembered influence judgements of learning (JOLs)? *Memory*, *18*(3), 351–362.

Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgements of retention (JORs). *The Quarterly Journal of Experimental Psychology*, *65*(7), 1376–1396.

Tekcan, A. I., Topçuoğlu, V., & Kaya, B. (2007). Memory and metamemory for semantic information in obsessive-compulsive disorder. *Behaviour Research and Therapy*, *45*(9), 2164–2172.

Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2011). Context affects feeling-of-knowing accuracy in younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 96–108.

Thomas, A. K., Bulevich, J. B., & Dubois, S. J. (2012). An analysis of the determinants of the feeling of knowing. *Consciousness and Cognition*, *21*(4), 1681–1694.

Thomas, J. P., & McFadyen, R. G. (1995). The confidence heuristic: A game-theoretic analysis. *Journal of Economic Psychology*, *16*(1), 97–113.

Thompson, W. B., & Mason, S. E. (1996). Instability of individual differences in the association between confidence judgments and memory performance. *Memory & Cognition*, *24*(2), 226–34.

Thomson, D. R., & Milliken, B. (2013). Contextual distinctiveness produces long-lasting priming of pop-out. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 202–15.

Tiede, H. L., & Leboe, J. P. (2009). Metamemory judgments and the benefits of repeated study: Improving recall predictions through the activation of appropriate knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 822–8.

Travis, S. L., Mattingley, J. B., & Dux, P. E. (2013). On the role of working memory in spatial contextual cueing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 208–19.

Treitz, F. H., Heyder, K., & Daum, I. (2007). Differential course of executive control changes during normal aging. *Aging, Neuropsychology, and Cognition*, *14*, 370–393.

Trosset, M. W., & Kaszniak, A. W. (1996). Measures of deficit unawareness for predicted performance experiments. *Journal of the International Neuropsychological Society*, *2*, 315–322.

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition*, *41*(3), 429–442.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organisation of Memory* (pp. 381–403). New York and London: Academic Press.

Tulving, E. (1983). *Elements of Episodic Memory*. Cambridge, UK: Oxford University Press.

Tulving, E. (1985a). How Many Memory Systems Are There ? *American Psychologist*, *40*(3), 385–398.

Tulving, E. (1985b). Memory and consciousness. *Canadian Psychology*, *26*(1), 1–12.

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, *21*, 437–484.

Tuna, S., Tekcan, A. I., & Topçuoğlu, V. (2005). Memory and metamemory in obsessive-compulsive disorder. *Behaviour Research and Therapy*, *43*(1), 15–27.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Vasterling, J. J., Seltzer, B., Foss, J. W., & Vanderbrook, V. (1995). Unawareness of deficit in Alzheimer's Disease: Domain-specific differences and disease correlates. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, *8*(1), 26–32.

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, *18*(2), 332–339.

Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review*, *105*(1), 33–57.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*, 277–95.

Widner, R. L., Smith, S. M., & Graziano, W. G. (1996). The effects of demand characteristics on the reporting of tip-of-the-tongue and feeling-of-knowing states. *American Journal of Psychology*, *109*(4), 525–538.

Widner, R. L., & Winkelman, S. E. (2005). Tip-of-the-tongue experiences are not merely strong feeling-of-knowing experiences. *The Journal of General Psychology*, *132*(4), 392–407.

Wilson, B. A., Cockburn, J., & Baddeley, A. D. (1985). *The Rivermead Behavioural Memory Test*. Bury St Edmunds, UK: Thames Valley Test Company.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176.

Woods, C. M., Vevea, J. L., Chambless, D. L., & Bayen, U. J. (2002). Are compulsive checkers impaired in memory? A meta-analytic review. *Clinical Psychology: Science and Practice*, *9*(4), 353–366.

Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(2), 187–205.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memroy judgments: A formal dual-process model and an analysis of Receiver Operating Characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1415–1434.

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.

Yonelinas, A. P., & Levy, B. J. (2002). Dissociating familiarity from recollection in human recognition memory: Different rates of forgetting over short retention intervals. *Psychonomic Bulletin & Review*, *9*(3), 575–582.

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, *133*(5), 800–32.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The Handbook of Ageing and Cognition* (2nd ed.). London: Lawrence Erlbaum Associates.

Appendices

Appendix A: Ethical approval

NHS ethical approval was granted by the Yorkshire and Humber Research Ethics Committee (ref: 11/H1308/14) for recruitment and testing of patients diagnosed with early stage dementia. The NHS approval covered recruitment of patients at a memory clinic and testing of patients within their own homes. All patients were recruited from a single memory clinic, Towngate House, under the direction of Dr Wendy Burn.

Separate institutional ethics was granted by the Institute of Psychological Sciences, University of Leeds. This covered the recruitment of young adults and healthy older adults. Young adults were recruited from the University of Leeds. Older adults were recruited through an existing Older Adult Panel held at the Institute.

Appendix B: One tailed t-tests to compare recognition against chance accuracy (0.25) for Experiment 3.3 and 3.4

Experiment 3.3

		M (SD)	
Young adults	2 cues	0.668 (0.14)	$t(40) = 19.320$, $p < .001$
	4 cues	0.698 (0.16)	$t(40) = 18.405$, $p < .001$
Older adults	2 cues	0.640 (0.19)	$t(19) = 9.225$, $p < .001$
	4 cues	0.703 (0.14)	$t(19) = 14.918$, $p < .001$

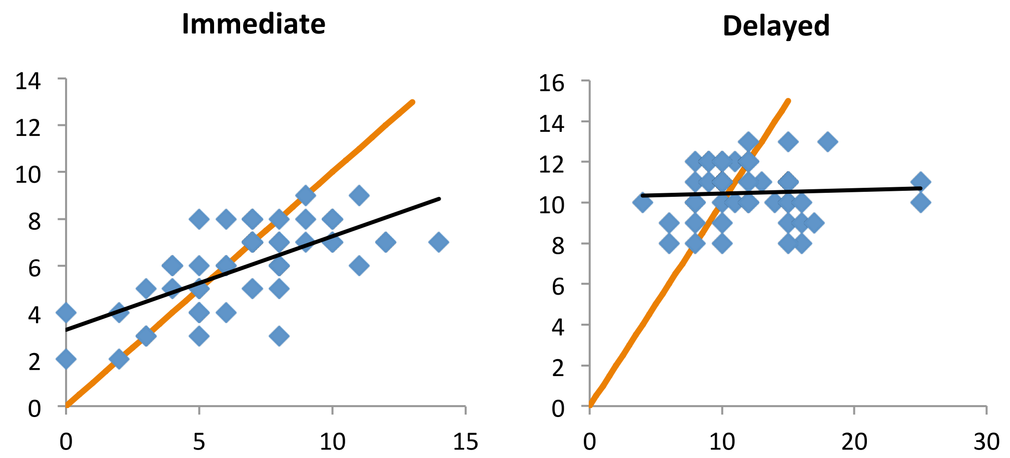
Experiment 3.4

		M (SD)	
Young adults	2 cues	0.472 (0.17)	$t(40) = 8.532$, $p < .001$
	4 cues	0.553 (0.16)	$t(40) = 12.065$, $p < .001$
Older adults	2 cues	0.303 (0.16)	$t(19) = 1.467$, $p = .051$
	4 cues	0.390 (0.18)	$t(19) = 3.503$, $p = .002$

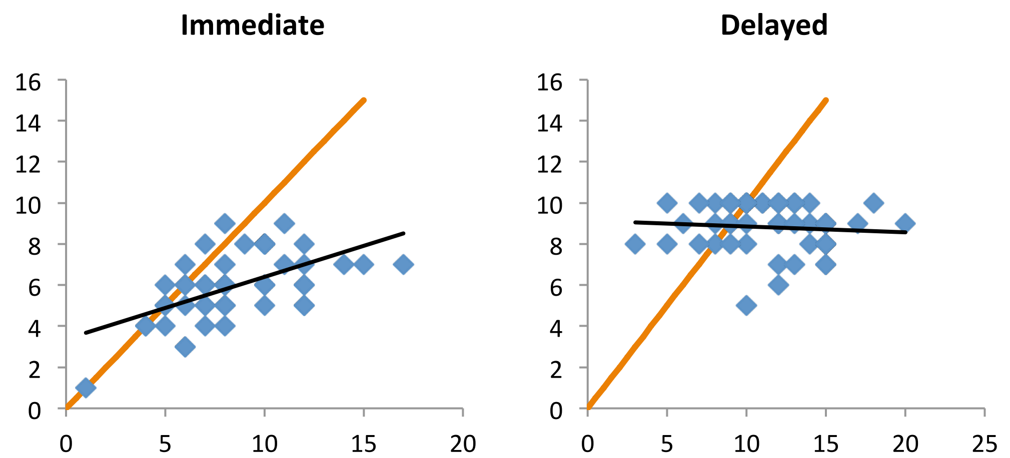
Appendix C: Scatterplots between JORF and retrieval failure for Experiment 5.1

Plots of JORF predictions (x-axis) against actual retrieval failure (y-axis). Orange trendline represents perfect accuracy. Black trendline represents observed data.

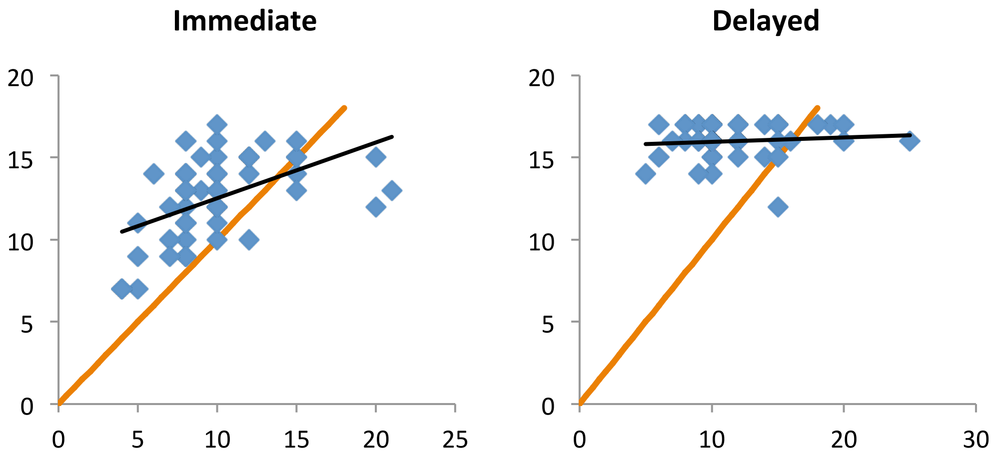
Fruits



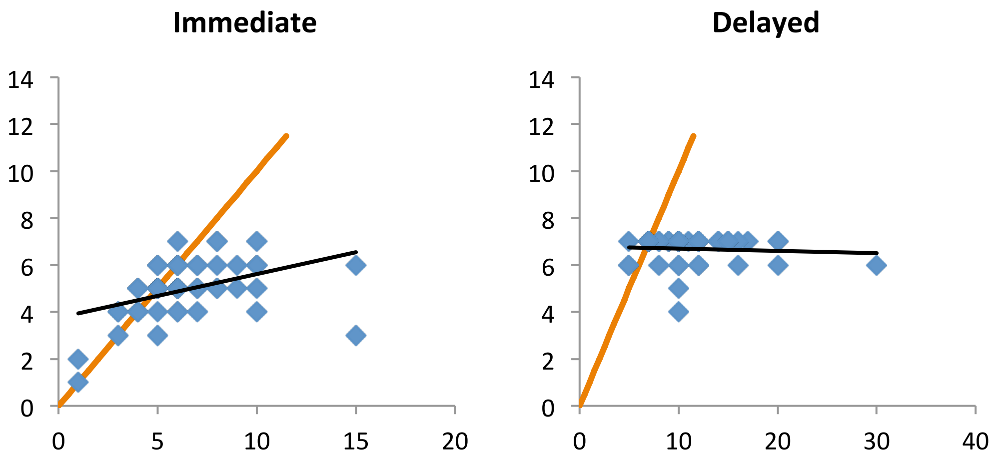
Colours



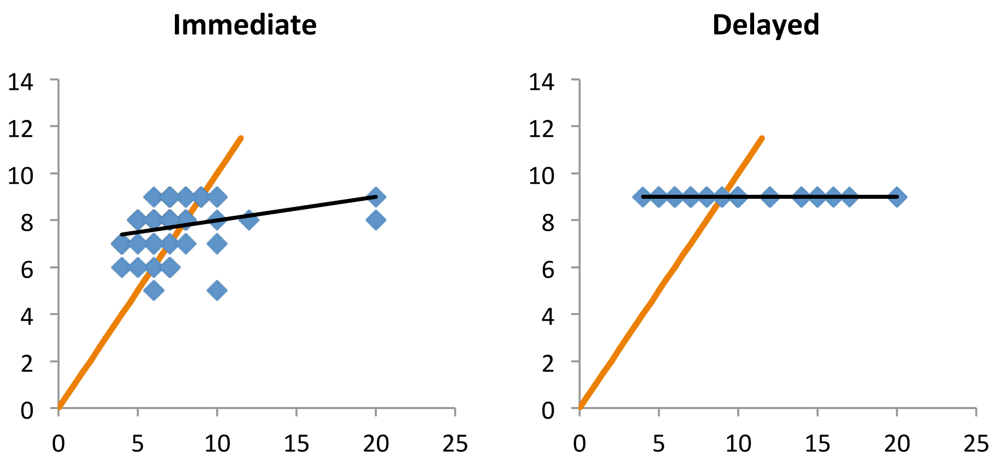
Creatures



French words



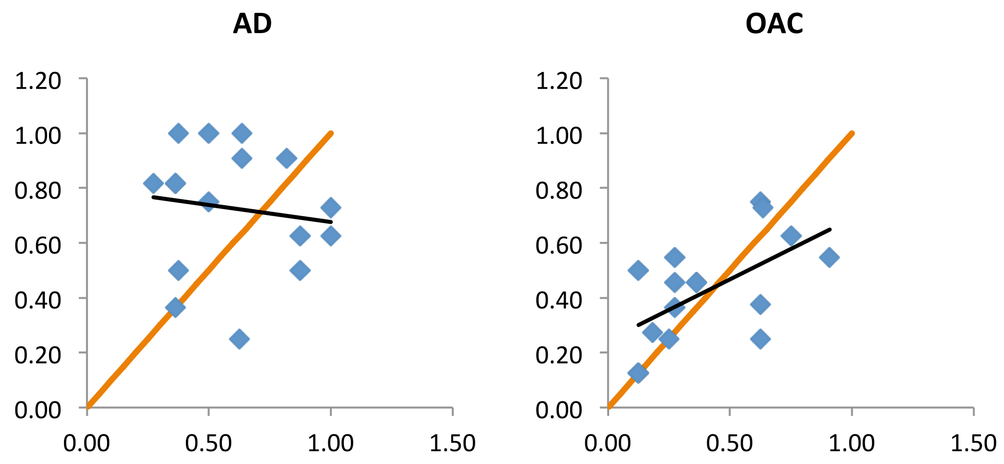
Nonwords



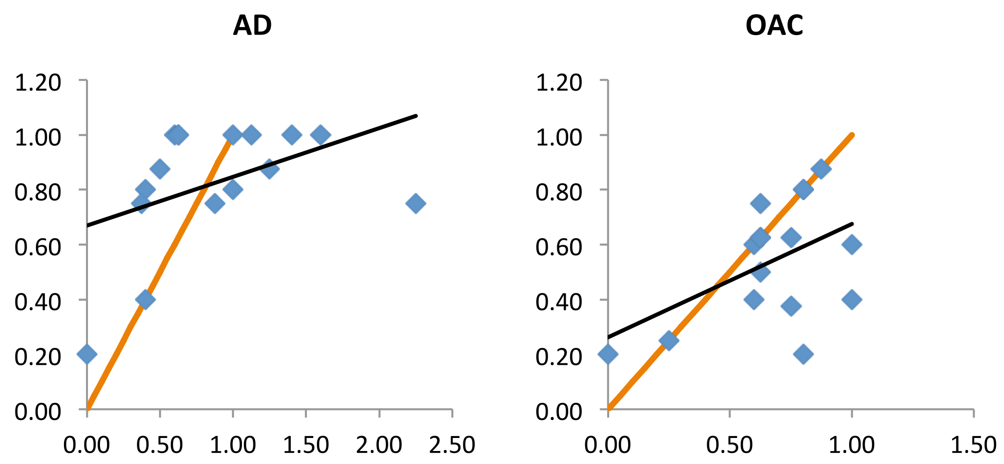
Appendix D: Scatterplots between JORF and retrieval failure for Experiment 5.3

Plots of JORF predictions (x-axis) against actual retrieval failure (y-axis). Orange trendline represents perfect accuracy. Black trendline represents observed data.

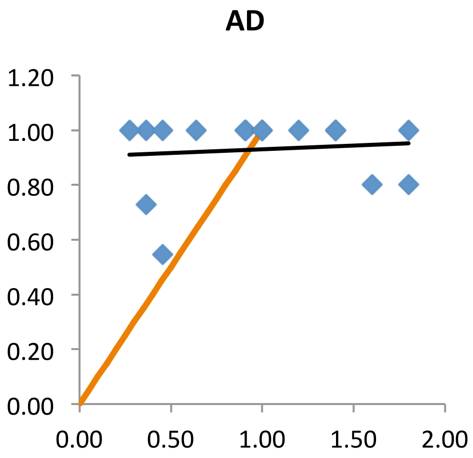
Fruits

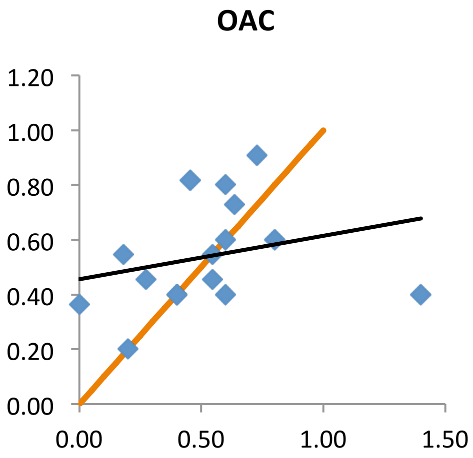


Colours

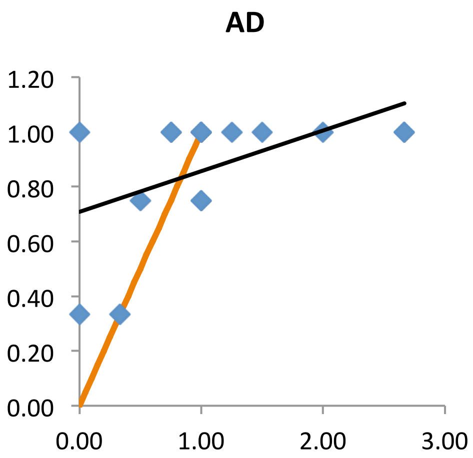


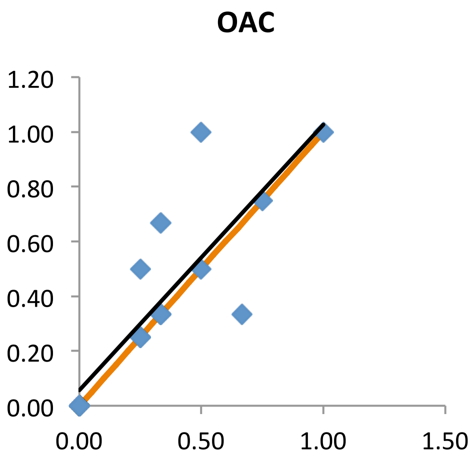
Creatures





French words





Nonwords