False Memories and Ageing:

Source-Monitoring Interventions Reduce False Recognition in Both Younger and Older Adults

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Clinical Psychology

At Massey University, Manawatu,

New Zealand

Rachael Sim

2015

Abstract

The purpose of the current research was to examine age-related differences in false recognition and attempt to establish whether these differences were best explained by the fuzzy-trace theory, source-monitoring processes (as part of the activation-monitoring theory), or sensitivity and/or criterion differences in signal detection ability. Eighty participants (40 younger adults, 16-30 years old, and 40 older adults, 75-80 years old) were randomly assigned to one of two experimental conditions. Twenty participants from each age group completed one of two versions of the Deese (1959) Roediger and McDermott (1995) false recognition task (DRM). The standard version required a simple old/new recognition judgement, while the source-monitoring version also required a source judgement. The results showed that older adults were sometimes, but not always, more prone to making false recognition errors compared to younger adults. Requiring source judgements decreased false recognition in both younger and older adults to a similar extent. Signal detection analyses showed that older adults were less sensitive than younger adults, and those in the source-monitoring condition were more conservative than those in the standard condition when making decisions about whether items were old. These and other results are discussed in terms of their implications and applications to real life false memories. As expected the results did not favour one theoretical perspective over another. Most of the results can be adequately explained by both the fuzzy-trace and activation-monitoring theory, although source-monitoring processes provided a simpler explanation of the research findings than fuzzy-trace theory or an appeal to bias and/or sensitivity differences.

This project was evaluated and approved by the Massey University Human Ethics Committee: Southern B (refer to Appendix A for approval letter).

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my main supervisor, Dr Stephen Hill, for the continuous support, enthusiasm, and guidance. His direction and vast knowledge aided me throughout the entire process of this thesis, and his encouragement provided perspective, often when it was well needed. I could not have imagined having a better advisor and mentor for this project.

I would like to thank my two secondary supervisors, Associate Professor John Podd and Dr Joanne Taylor. In particular, I thank John for his expertise and guidance in relation to signal detection theory, and I thank Jo for her knowledge and supervision regarding the clinical aspects of my doctoral degree. Both John and Jo's expertise, guidance, and support played a vital role in completing my research. Also, I thank Mr Malcolm Loudon, whose technical skills greatly helped with my signal detection results.

I am also grateful for the financial assistance provided to me by Massey University (Doctoral Scholarship and Purehuroa Award for Māori Postgraduate Students), the Ministry of Health Haoura Māori Scholarship, the HOPE Foundation for Research on Ageing, and education grants from the Taranaki Māori Trust Board and Te Ātiawa o Te Waka-a-Māui Trust. It has meant a great deal that these organisations have supported me and my chosen areas of research and learning.

To my family and friends I am extremely thankful. As they know, over the last four years I've described the journey of writing my thesis in many ways (few of them good), and under no circumstances could I have started and completed this journey without them. My partner, James Montgomery, has been my right-hand man. Numerous times he pushed me back into the fight when I felt like giving up, and he always knew what to say (or when to say nothing at all) to lift my spirits and keep me balanced. Nothing I write here can convey how much it has meant to me to have had James's support and love throughout this task.

My family, although rarely seen, where always with me and each provided support in their own unique way. Thanks Fran and Rory for all the delicious edibles, without your preserves my fridge would often have been completely empty. Serah, thanks for the long-distance phone calls and having something other than study to chat about. A very special thank you goes to my parents (Anna and Kevin Sim), I am extremely grateful for all of the support and encouragement you have provided me during my entire university life.

Lastly, I would like to take a moment to appreciate the other doctoral students I have come across and got to know on my journey. I will not name you all, you know who you are. You are the ones who provided me with some light relief, laughter, and fun. Those who I will name, Amber and Karen, took this journey with me, and my experience was greatly enhanced by being able to share it with these two who were walking the same road at the same time.

Table of Contents

List of Tables

List of Figures

Chapter One

Introduction

As we get older a number of our memory abilities decline (Bäckman, Small, Wahlin, & Larsson, 2000; Park, 2000; Park & Minear, 2004). This deterioration not only includes a decrease in abilities, but also an increase in the number of memory errors that we make (e.g., G. Cohen & Faulkner, 1989; Dywan & Jacoby, 1990; Law, Hawkins, & Craik, 1998; Lövdén, 2003). When people remember, two types of errors can occur: errors of omission and errors of commission. An error of omission is when an individual attempts to retrieve a memory, but fails to do so. An error of commission occurs when an individual remembers an event somewhat differently from the way it happened, or when an individual remembers an event that never happened (Roediger & McDermott, 2000a). The former, forgetting, is well known, well researched, and everyone has experienced it. Errors of commission are more controversial. One type of error of commission of particular interest is false memories: remembering or believing an event occurred when it did not. Everyone would prefer to believe that their memories are accurate representations of the past and fortunately, memory does operate rather accurately across numerous conditions and situations (Schacter, 1995). However, everyday experience and laboratory demonstrations have shown that memory distortions and false memories can and do occur, and that under certain circumstances memory can be surprisingly inaccurate. Such events have led to memory illusions and false memories now holding a prominent position in contemporary cognitive psychology as well as being a popular topic in the public domain.

Great effort has been spent attempting to establish the underlying mechanisms that facilitate accurate remembering, and from this effort two widely endorsed theories have been established to explain the creation of false memories and age-related differences in false memory tasks. They are the fuzzy-trace theory (Reyna & Brainerd, 1995b) and the activation-monitoring theory (Roediger & McDermott, 1995). Curiously, neither of these theoretical frameworks can best account for the creation of, or age-related increases in, false memories. These two theories are often set against one another, but it is difficult to design an experiment that can decisively support one theory over the other. This is because there is substantial overlap in the claims made by the two theories. Even if researchers favour one of the theories, they will often discuss their experimental results using both. A third theoretical approach, signal detection (Macmillan & Creelman, 2005), has been used to help explain false recognition findings, including age-related differences (of which, more later).

The general consensus is that older adults perform more poorly in false memory tasks than their younger counterparts, because many researchers have found age-related deficits in the memory processes believed to underpin false memories. One of these memory processes is source-monitoring (Hashtroudi, Johnson, & Chrosniak, 1989; Mitchell, Johnson, & Mather, 2003), which constitutes half of the activation-monitoring theory (the other half is semantic activation). Research has also shown older adults have problems using gist and verbatim memory traces – the mechanisms in the fuzzy-trace theory – effectively (Budson, Daffner, Desikan, & Schacter, 2000; Koutstaal & Schacter, 1997; Pierce, Sullivan, Schacter, & Budson, 2005). Research into age-related differences in false memories has typically found mixed results, with older individuals performing more poorly in some studies and no differently from younger people in others. These mixed findings have raised a number of questions about false memories and the mechanisms behind them that are still in the early stages of being explored.

The most commonly employed method (and the one used in the current research) to elicit false memories is the Deese (1959) Roediger and McDermott (1995) paradigm, commonly known as the DRM paradigm. In this paradigm participants are presented with lists (or a list) of semantically related words, and each list primes a common associate word – known as the critical lure – that was not presented during learning. One example of a DRM list is: *woman, husband, uncle, lady, mouse, male, father, strong, friend, beard, person, handsome, muscle, suit, old.* After learning this list participants will likely falsely recall and/or recognise the word *man*. Each DRM list was created using the 1952 Minnesota norms for the Kent-Rosanoff word association test (Jenkins, 1970). The DRM paradigm is a robust false memory phenomenon, and DRM lists can be presented and tested numerous ways to reliably cause false remembering.

The purpose of the current research is to examine the age-related differences in a DRM false recognition task and attempt to establish if age-related differences in false recognition are best explained by (a) fuzzy-trace theory, (b) source-monitoring processes, or (c) criterion and/or sensitivity differences in signal detection ability. Finding results that favour one mechanism over another is a difficult task. It would take a specific pattern of results to support one but not another theoretical perspective, and such an event is unlikely. It is more likely that specific results will partially favour one theory, while other results will partially favour the opposing theory.

The current research is noteworthy because in an attempt to fulfil its purpose it will employ a novel combination of procedures and analyses, and by doing so will deepen our current understanding into the creation of false memories and age-related effects. In the present research the well-used DRM paradigm will be employed alongside a source-monitoring manipulation shown to improve recognition performance (Multhaup & Conner, 2002). This is important because many DRM tasks that incorporate a source-monitoring aspect have been found to *decrease* performance (by increasing false recognition) in younger and older adults, a finding which scholars have difficulty explaining given that source-monitoring is theorised to improve performance (Hicks & Marsh, 1999).

Source memory is critical because it is necessary for a number of cognitive tasks. In experimental research examining memory, source memory allows participants to discriminate test items they recall or recognise as studied items, from test items that are familiar but originate from sources other than the studied items. In daily life, source memory adds to one's ability to exercise control over one's opinions and beliefs. To illustrate, if you are presented with a fact you feel you have come across before, and remember it to be from either a reliable (e.g., a textbook or national geographic magazine) or a less reliable (e.g., a tabloid magazine or internet site) source, you have the information necessary to evaluate the fact's veracity. The sensation of remembering a particular experience from your past depends on source acknowledgements based on specific phenomenal characteristics of the remembered experience. Without such phenomenal qualities, the information is not experienced as memory, but as knowledge or values (M. K. Johnson, Hashtroudi, & Lindsay, 1993).

Additionally, the recognition test used in the current research will include weak lure words (lure words that were not presented during the study phase but are weakly semantically related to their corresponding DRM list)[1]. Weak lures are believed to be a more

---

[1] The information about the semantic relationship between words and word lists is based on the 1952 Minnesota norms for the Kent-Rosanoff word association test (reprinted by Jenkins, 1970). For the example DRM list provided above three weak lures could be: *boy, dog, pants*.

sensitive measure of false recognition, and are expected to provide more information about age-related false memory differences than if only critical lures were examined (Tun, Wingfield, Rosen, & Blanchard, 1998). For reference, Table 1.1 below provides a glossary of DRM test items employed in the current research.

The novel analyses that will be employed are signal detection analyses (as well as commonly used basic recognition analyses). Signal detection theory has been underutilised in false memory research due to its controversial start in the area. However signal detection is useful for understanding false memories, because it provides information about how participants make recognition decisions. Unlike previous research, in which signal detection findings have been given a subordinate position, in the present study, signal detection ability will be examined and discussed alongside the other theoretical approaches. Another significant factor in the present research is that the sample of older adults will be considerably older (75-80 years) than the younger adults (16-30 years). This is significant because research often shows older adults aged (approximately) 60-70 years perform similarly to younger adults on some cognitive tasks, whereas significant decreases in performance appear during later decades (Bäckman & Larsson, 1992; Colsher & Wallace, 1991; Korten et al., 1997).

Table 1.1

Glossary of DRM Test Items

Critical lures	Critical words not presented during study that are strongly semantically related to their corresponding DRM list
Weak lures	A type of critical lure. Weak lures are words not presented during study that are weakly semantically related to their corresponding DRM list
Unrelated items/words	Words not presented during study that are not semantically associated with DRM word lists that were presented
Presented items/words	DRM list items that were presented during study

In terms of the limits of this study, it will be useful here to illustrate the differences between false, repressed, recovered, and discovered memories, as the present research only covers false memories. Although theoretically the four phenomena are connected (Conway, 1997), there is little empirical research to substantiate the connection (Kihlstrom, 2004). Known as the recovered memory debate, the effortlessness in which individuals can be directed to remember an event that did not happen and believe the memory with high confidence has caused controversy that crosses the boundaries of both clinical and experimental psychology (Neath & Surprenant, 2003).

The notion of repressed memories came from the clinical work of Sigmund Freud (1896). Repressed memories were defined as unpleasant memories that were being consciously avoided through repression, the unconscious defence mechanism used to exclude painful memories from consciousness (Erdelyi, 1985). Freud held that repressed memories (specifically of abuse and trauma) caused persisting pathological symptoms and thus the aim of psychoanalysis was to allow clients to retrieve these memories. However, Freud soon abandoned this perspective, as he realised many of the ‘memories’ clients retrieved were confabulations (Schacter, 1995). Despite this abandonment, Freud’s ideas have been extremely influential in psychology and they continue to cause controversy.

It is important to understand repressed memories because the idea that memories can be recovered inevitably means that memories, in some way, do become inaccessible for a length of time (Conway, 1997). One problem with the term *recovered memories* is the implication that the memory *is* of a true event (Neath & Surprenant, 2003). Therefore, the term *discovered memories*, which is neutral in regards to the memory being real or not, has been applied (Schooler, Bendiksen, & Ambadar, 1997). Using the notion of a discovered memory, a recovered memory requires three distinct and independent characteristics. First, there needs to be evidence that the memory is consistent with a real event. Second, there must be a period of time during which the individual did not remember such an event. Third, the discovery of the experience must be authentic; the individual must have a profound sense that they have discovered an event in their past that they genuinely believe they had no knowledge of (Schooler et al., 1997). For example, there have been cases, although few in number, of childhood sexual abuse that have been objectively demonstrated to have previously occurred, along with a length of time in which the individual (now an adult) did not remember the abuse, and the memory of the abuse was genuinely discovered (Schooler et al., 1997). Although the recovered memory debate is interesting, it is beyond the scope of the present research so is not discussed further.

To achieve the aim (stated above) of the current research, Chapter 2 provides an overview of the cognitive changes we experience as we age, with particular attention given to source memory abilities. The two theories of false memories and their overlapping characteristics are also discussed thoroughly. Additionally, throughout Chapter 2 a variety of results from research into memory distortions and false memories are considered. Chapter 3 focuses on the use of signal detection theory in false recognition research. Based on the information gathered in Chapter 2 and 3, Chapter 4 provides the research method developed and employed to examine age-related differences in DRM recognition in the current research. Chapter 4 highlights the specific tasks and measures that provide the ability to examine which theoretical perspective the results favour. Chapters 5 and 6 encompass the results, analysing the findings using common recognition accuracy analyses, and signal detection analyses. Lastly, in Chapter 7 the age-related results are discussed in terms of the three theoretical perspectives and conclusions are drawn regarding which of the perspectives best explains the age-related findings.

Chapter Two

Literature Review Part One: False Memories

Memory is one of the most important abilities we have, and everyone would prefer that their memories correctly reflected their past. However, everyday experience and research has demonstrated this critical tool is not errorless. A false memory is one type of error that our memory can create, and this type of error has been shown to increase as we get older. The current literature review will address a variety of areas related to false memory research. In the first section, the cognitive changes that occur as we age will be outlined, followed by a thorough discussion of age-related changes in source memory abilities and the possible underlying causes. Then background information about false memories and memory distortions is provided, along with a review of research that has examined ageing and performance on false memory tasks. The next section discusses the fuzzy-trace theory and the activation-monitoring theory, with a particular focus on the features common to both. The second part of the literature review covers signal detection theory and false recognition research that employs signal detection. This last section begins with an overview of signal detection theory and age-related differences in signal detection estimates for recognition tasks. Then the two proposed models (criterion-shift model and storage-based model) of false recognition based on signal detection results are discussed, followed lastly by a review of DRM (Deese/Roediger-McDermott) research in which signal detection analyses were used to examine findings.

Cognitive Ageing

We have all heard older adults say in one way or another that their memory is failing them because they are getting older. Most of the comments people make about their believed declining cognitive function, such as ‘I’m not as sharp as I used to be’, reflect a socially-shared metaphor concerning how we require cognitive resources (i.e., a supply of mental power) to manipulate information and solve problems, and that as we age this resource somehow declines. This view is referred to as a resource model of ageing (Park, 2000). Of central concern to this mental resource is whether there is an underlying mechanism that can explain age-related cognitive decline. This section will first provide an overview of the changes in performance on common cognitive tasks across the life span, before discussing the main mechanisms hypothesised to explain age-related cognitive decline.

A remarkable observation from research into cognitive ageing is that it appears many forms of memory are negatively impacted by age. Park and Minear (2004) gathered data from a lifespan sample of 345 adults aged 20 to 95 years old. The sample were tested on a number of tasks that tapped processing speed (the speed at which cognitive operations occur), working memory, long-term memory, short-term memory, and knowledge-based verbal ability. Their results demonstrated a steady and continuous decline in working memory, short-term memory, long-term memory, and processing speed, as age increased, with approximately equal decline in the four mechanisms across the decades. By contrast knowledge-based verbal ability remained relatively stable, showing small increases with age. Similar results have been reported by Salthouse (2010). Age-related cognitive deficits have also been observed in tasks examining episodic memory (Craik & Jennings, 1992; Kausler, 1994), implicit memory (Hultsch, Masson, & Small, 1991), and semantic memory (Crook & West, 1990).

In a chapter titled *Cognitive functioning in very old age,* Bäckman et al. (2000) discussed the problems with generalising research that involves individuals considered young-old (55 to 74 years old) to those considered old-old (75 to 85 years), and oldest-old (85 years and older). The important question is whether cognitive decline in very old age is gradual, accelerated, or attenuated, compared to the cognitive decline in younger adults. Notably, the definitions (young-, old-, oldest-old) are arbitrary as the categories are not stable entities – they change across time and cultures, especially with regard to differences in life-expectancy. In Western societies those considered young-old, old-old, and oldest-old, characterise the fastest growing segment of the population. According to Statistics New Zealand (2000), by 2051 the projected growth of adults 65 and older is expected to be 166%, compared to the base population in 1996. With one out of four New Zealanders being 65 or older and there will be approximately 60% more older adults than children under 15 years of age. The current life-expectancy for males is 74 years and for females 80 years. By 2051 these numbers are expected to increase to 80 years for males, and 84.5 years for females.

With such information about the growth and ageing of the older population it is surprisingly difficult to find cognitive research examining aspects of memory in groups in late senescence. That which does exist comes mainly from large-scale multidisciplinary research, in which many individual characteristics are measured. Colsher and Wallace (1991) conducted longitudinal research with community-dwelling older adults (three groups aged 65-74, 75-84, and 85 onwards). They compared the three age groups on an altered version of the Short Mental Status Questionnaire (Pfeiffer, 1975), a recall task, and a recognition task, recording a baseline measure, and 3- and 6-year follow-ups. Overall, their results demonstrated older age was associated with poorer performance, as well as a greater decline in performance, for all three tasks.

Hultsch, Hertzog, Small, McDonald-Miszczak, and Dixon (1992) examined changes in episodic memory in older adults (aged 55-70 and 71-86) by conducting 3-year follow-up research. They found that episodic memory performance declined only in the older group. At the 3-year follow-up the older group's performance had declined by almost half a standard deviation. The 55-70 year olds had declined by less than .20 standard deviations. Further evidence of accelerated decline in old-old groups compared to young-old groups on episodic memory tasks has been observed by Bäckman and Larsson (1992), Crook and Larrabee (1992), Korten et al. (1997), and Lamont (2006). Also, a study which measured performance on a number of working memory measures across 19 to 96-year olds demonstrated a systematic decline with ageing. Small decrements in overall working memory performance were found with increasing age up until 80 years old. After 80 years, the decline in performance across the decades was much greater compared to the previous decades (Gilinsky & Judd, 1994).

Many researchers who investigate cognitive ageing believe it is the decline of a single mechanism (Salthouse, 1996), or a limited number of mechanisms (Verhaeghen, 2011), that produces widespread decline across cognitive functions, and substantial effort has been invested to examine this belief. Salthouse (1996) has proposed a well-developed theory in which the fundamental mechanism accounting for age-related variance in memory performance is a generalised decrease in the speed of which individuals can perform cognitive operations (i.e., processing speed). Salthouse hypothesised that two important functions account for the relationship between processing speed and cognition, the limited time mechanism, and the simultaneity mechanism. The limited time mechanism is how "the time to perform later operations is greatly restricted when a large proportion of the available time is occupied by the execution of early operations" (p. 404). The simultaneity mechanism is how "the products of early processing may be lost by the time that later processing is completed” (p. 405). Older adults perform more poorly on cognitive tasks compared to younger adults, because older adults are slow to execute early steps in a task, and this slowness can mean they do not reach the later steps, and that the earlier steps are not available to them when they are needed. Salthouse assembled a number of results from research conducted by himself and others demonstrating that a large proportion of age-related variance in a number of cognitive tasks (from memory to reasoning) can be explained via the performance on perceptual speed tasks (simple paper-and-pencil measures which require one to make quick same-different judgements about pairs of digits, letters, or symbols). Other research supports this position, demonstrating that processing speed accounts for a considerable portion of the variance in the performance of episodic memory tasks; when statistical methods are used to control processing speed age-related variation notably reduces (e.g., Park et al., 1996; Salthouse & Babcock, 1991; Verhaeghen, 2011).

Another view is that a decline in working memory capacity causes decreased cognitive performance with ageing (Craik & Byrd, 1982). Working memory is the entire pool of ‘mental energy’ available to a person, to store, retrieve, process, and manipulate information. Capacity refers to the space available to briefly store information and/or the space available for processing task-relevant information (Baddeley, 1986). Ageing may result in reduced storage capacity, reduced processing capacity, or both (Kausler, 1994). There is substantial evidence illustrating performance on working memory tasks declines with age. Park et al. (1996) examined a life span sample of 301 adults (aged 20-90 years old) and their performance on multiple indices of working memory (backward digit span task, a computational span task, and a reading span task). Their results demonstrated a systematic decline in working memory performance across the life span. Comparable results have also been found by a number of other researchers (e.g., Dobbs & Rule, 1989; Gick, Craik, & Morris, 1988; Salthouse & Babcock, 1991; Salthouse, Kausler, & Saults, 1988). Meta-analyses by R. E. Johnson (2003) examined 16 articles that investigated working memory span, ageing, and learning and memory of texts. Results showed a statistically significant age deficit in working memory, with a mean weighted effect size of .78. Meta-analyses by Verhaeghen, Marcoen, and Goossens (1993) analysed 17 studies on ageing and working memory span. Results found a significant negative relationship between age and working memory, with a mean weighted effect size of .81.

Some research has failed to find age-related declines in working memory capacity (Hartley, 1986, 1993). Moreover, age-related differences in performance on other memory tasks (i.e., not working memory tasks) are not always attributable to differences in working-memory performance (Hartley, 1986, 1993; Light, 1991; Verhaeghen, 2011). There is also little consensus as to which specific aspect of working memory declines with age (Craik & Jennings, 1992; Salthouse, 1990). There is evidence to suggest working memory capacity is important for cognitively controlling false memories; that is, the ability to identify a critical lure as such, and correctly respond that it was not presented during study (McCabe & Smith, 2002; Watson, Bunting, Poole, & Conway, 2005). Individuals with a high working memory capacity have been found to have lower levels of false remembering than those with a low working memory capacity (Leding, 2012), and working memory capacity has been found to correlate negatively with false recall (Unsworth & Brewer, 2010) and false recognition (Peters, Jelicic, Verbeek, & Merckelbach, 2007).

Hasher and Zacks (1988) argue that working memory decline in older adults is not a decline in working memory capacity, but rather a breakdown, in the inhibitory processes that control the contents of an individual’s working memory. The inhibition deficit could be for either, irrelevant external stimuli (i.e., information that is irrelevant with respect to what needs to be remembered), or irrelevant internal stimuli (i.e., thoughts that are divergent from the current memory task; Kausler, 1994). In either case, because older adults fail to inhibit irrelevant stimuli, their working memories become filled with irrelevant material, which places increased strain on their capacity to process relevant information.

Research into the efficiency of inhibition indicates older adults have a specific deficit in inhibitory processes compared to younger adults (Rowe, Valderrama, Hasher, & Lenartowicz, 2006). Schelstraete and Hupet (2002) and West and Alain (2000) observed age-related deficits in the classic inhibitory Stroop task (Stroop, 1935). Age-related deficits have been found in negative priming tasks (Hasher, Stoltzfus, Zacks, & Rypma, 1991; May, Kane, & Hasher, 1995). Borella, Carretti, and De Beni (2008) and Persad, Abeles, Zacks, and Denburg (2002) demonstrated a steep decline in inhibitory performance occurs after age 60. Interestingly, research has revealed individuals who performed poorly on the Stroop task, regardless of age, are more prone to DRM false memories (Alberts, 2007). Similarly, heightened seriation in the Random Number Generation task (Ginsburg & Karpiuk, 1994) is associated with high false recognition rates (Peters, Jelicic, Haas, & Merckelbach, 2006).

Lövdén (2003) reported significant correlations between false recall and memory tests believed to reflect inhibition. However, using structural equation modelling, Lövdén demonstrated that inhibition only had an indirect effect on age-related false memories via episodic memory performance. Overall, these results suggest that inhibitory processes play some role in false memory performance. This role may be indirect and related to the processes involved in monitoring false memories (as will be discussed below).

Executive control is another mechanism thought to underpin the widespread age-related declines in cognitive tasks. Executive control is a key aspect of working memory which encompasses inhibition. Loosely defined, executive control is a general purpose mechanism that controls and regulates cognitive operations (Miyake et al., 2000). Executive control is not a unitary construct, research suggests that it can be divided into a minimum of four distinct, but related, characteristics (Engle, Tuholski, Laughlin, & Conway, 1999; Miyake et al., 2000). These characteristics are (a) resistance to interference, also named inhibition, the ability to intentionally inhibit automatic and/or dominant responses when the task requires; (b) the ability to coordinate distinct tasks, which is mainly measured using dual-task paradigms; (c) task shifting, the ability to switch between tasks or mental sets; and (d) updating and monitoring representations in working memory, the ability to monitor and code incoming information for its relevance to the task being done, then revising items in working memory by replacing old irrelevant information with new relevant information (Miyake et al., 2000; Verhaeghen, 2011).

Verhaeghen (2011) conducted a meta-analysis and a structural equation model on the elements of executive control described above and ageing. The results of the meta-analysis indicated that age-related deficits are, generally, found in tasks that require executive control compared to tasks that require only minimal executive control. The correlation matrix of Verhaeghen’s structural equation model included age, three underlying aspects of cognition (processing speed, short-term memory, and working memory), three complex cognitive abilities (episodic memory, reasoning ability, and spatial ability), and two aspects of executive control (resistance to interference and task shifting). The model indicated that executive control did not explain any of the age-related variance in the complex cognitive tasks, or any of the other abilities measured. In the model age-related variance in executive control was fully explained by age-related differences in processing speed, and processing speed mediated age-related differences in both complex cognition and combined short-term and working memory. Verhaeghen's results indicated that age-related decline in executive control is not universal and executive control does not play an important role in more complex cognition. However, it appears the basic mechanism of processing speed does play an important part in age-related declines in complex cognitive tasks.

Overall, it seems that there are many ideas about what underlying mechanism is responsible for age-related decline in memory, but none of these are overwhelmingly supported by research findings. The following section focuses specifically on source memory abilities and ageing. Source memory is important because it is related to one's ability to differentiate sources in false memory tasks, and as will be seen older adults have a widespread source memory deficit.

Source Memory

'Source' can be defined as the characteristics that, jointly, detail the circumstances in which a memory is attained; for example, when and where the event transpired, the event's social context, and the means and modalities through which it was experienced (M. K. Johnson et al., 1993). Source memory is a memory for the origin of information encoded in a memory (M. K. Johnson et al., 1993). Put simply, source memory involves recollection of the origin of an item or information (Schacter, Osowiecki, Kaszniak, Kihlstrom, & Valdiserri, 1994). Source information is important because it allows one to establish why an item feels familiar (Dodson & Schacter, 2002).

A process believed to be important in false memory creation is *source-monitoring* (Gallo, Roberts, & Seamon, 1997; M. K. Johnson et al., 1993; Norman & Schacter, 1997; Roediger & McDermott, 1995; Roediger, Watson, McDermott, & Gallo, 2001; Skinner & Fernandes, 2009). Source-monitoring is the processes required to attribute memories, knowledge, and beliefs to their origins (Hashtroudi et al., 1989). The notion of source-monitoring is derived from the reality monitoring framework (M. K. Johnson & Raye, 1981). Reality monitoring concerns the segregation of information that is generated internally (e.g., imagined) from that which originates from an external source (e.g., observed events. This is referred to as internal-external discrimination). Source-monitoring involves internal-external source-monitoring, as well as external source-monitoring; differentiating external sources (e.g., did Jane or Simon say that?), and internal source-monitoring; differentiating internal sources (e.g., was that a memory of something I thought about or something I said?; M. K. Johnson et al., 1993).

To make a source judgement we rely on internal (cognitive operations that were created during encoding; e.g., records of elaboration, organisation, retrieval, and identification) and external (perceptual details, the context of the memory, semantic details, and emotion details) characteristics of memories *and* judgement processes (M. K. Johnson et al., 1993). The judgement processes require source decisions to be made based on (a) the differences between internal and/or external memory characteristics of sources and (b) the match between these memory characteristics and schemas that are consistent with a specific source. An example of the former is that for a real event one would have access to more external characteristics. Memories created internally (e.g., creating an imaginary picture) will be associated with internal memory characteristics (M. K. Johnson et al., 1993; Roediger, Watson, et al., 2001). By contrast, an example of the latter, when the aural characteristics for a memory of a fact matched one's schema for the voice of a specific friend, the fact would then be attributed to that friend (M. K. Johnson et al., 1993). During retrieval, memories that carry experiential details are often endorsed as true past occurrences, even though this endorsement may be mistaken (Jacoby, Kelley, Brown, & Jasechko, 1989). As source-monitoring is a mechanism implicated in false remembering it will be discussed thoroughly below, but first age-related differences in source memory and source-monitoring will be examined.

Source memory and ageing. Various research findings reveal that older adults have more trouble remembering source information compared to younger adults. For instance, McIntyre and Craik (1987) presented a series of questions and answers about Canada, either visually or aurally (two external sources), to older (mean age 69.2) and younger (mean age 19.4) adults. At test the participants were given a series of old and new questions about Canada to answer, and had to state the source of their answer. Sources included the experimental sources and sources from outside the experiment. The research showed that older adults had poorer memory for the source of their answers relative to the younger adults. Research has also demonstrated older adults have poorer source memory compared to younger adults of words (Ferguson, Hashtroudi, & Johnson, 1992) and fictitious statements (Schacter, Kaszniak, Kihlstrom, & Valdiserri, 1991).

To investigate age-related source memory differences using the reality monitoring paradigm, Hashtroudi, Johnson, and Chrosniak (1989) had participants say, think, and listen to words. Also, G. Cohen and Faulkner (1989) had participants imagine, watch, or perform different actions. Both of these experiments found older adults (mean age 69.4 and 76 years, respectively) had more difficulty correctly identifying the original source of the words or actions than younger adults (mean age 19.5 and 31, respectively). Likewise, Henkel, Johnson, and De Leonardis (1998) found older individuals (mean age 74.2 years) judged items that had been imagined as items that had been seen, more often than younger adults (mean age 20.2 years). Rosa and Gutchess (2011) found older adults (aged 61-91) had more trouble, compared to younger adults (aged 18-26), correctly remembering the source of actions that were performed by either themselves or another person.

In the above studies older individuals not only showed source memory deficits, but also deficits in recall and/or recognition performance. These findings raised the question of whether the demonstrated source memory declines simply reflected a general memory decline (Schacter, Koutstaal, & Norman, 1997). To answer this question researchers have examined if older adults’ source memory deficits were disproportionate to their recall and/or recognition deficits. Schacter et al. (1991) found that when item memory for fictitious statements was equal in older (mean age 69) and younger (mean age 19.3) participants, the older adults had poorer source memory for which of two experimenters, one female and one male, presented the item. Similarly, Ferguson et al. (1992) found that when older (mean age 69.8) and younger (mean age 20) adults’ recognition of words was equivalent, older adults exhibited a larger source memory deficit (sources were two female speakers).

McIntyre and Craik (1987) conducted a second experiment in which participants learnt fictitious statements about famous and fictional individuals. The facts were presented aurally or visually to participants, whom were later tested on their source and item memory. Again, older adults (mean age 69.7) exhibited poorer item and source memory in comparison to younger adults (mean age 23.3). When source recall was conditional on recalling the correct item, older adults made more source errors than younger adults. In particular, older participants displayed a predisposition to incorrectly identify sources outside of the experiment for statements that could only have been learnt in the experimental condition.

McDaniel, Lyle, Butler, and Dornburg (2008) examined age-related differences in reality monitoring when source judgements were conditional on correct recognition. The results showed that older adults (mean age 76.3) made more source errors, often stating actions that were only imagined were imagined *and* performed, compared to younger adults (mean age 19.0). Spencer and Raz (1995) conducted a meta-analysis to investigate age-related differences in memory for content (recognition memory) and context (source memory). The meta-analysis included 46 studies and found age-related differences for both content (effect size .58) and context (effect size .87) memory, with age-related source memory deficits being disproportionally greater than recognition memory deficits.

Schacter et al. (1994) considered the possibility that source memory deficits in older adults were due to difficulty with tasks when a small number of sources (generally two) are associated with a large number of items (many-to-one source mapping), because most source memory investigations employ this paradigm. Schacter et al. argued that, older individuals could be sensitive to the effects of interference that occur when a cue, in this case a source, is overloaded by being linked to a large number of items. The foundation for this idea is the cue-overload hypothesis, in which the likelihood that an item is recalled decreases as the number of items that are under a specific retrieval cue increases (Watkins & Watkins, 1975). However, Schacter et al. found contrary evidence when examining the cue-overload hypothesis. Older adults (mean age 68.3) had greater source deficits than younger adults (mean age 19.8) when both one-to-one (each item had a different source) and many-to-one source mapping were examined. The above evidence suggests that there are age-related source memory deficits independent of other memory difficulties (although age-related source memory deficits may be due to an age-related decline in a yet un-established cognitive mechanism), and these age-related source memory deficits likely decrease performance in false memory tasks.

Proposed causes of age-related source memory deficits. With the substantial amount of research reporting age-related deficits in source memory abilities it is surprising that there is no general consensus of the underlying cause of the age differences observed. As stated above, source memory depends on the characteristics associated with a memory and judgement/decision processes, both of which have been, independently, implicated as the reason older adults have poorer source memory than younger adults. Though the accounts for age-related impairment in source memory vary, two overarching themes have been identified, and both depend on the characteristics associated with a memory. The first theme is that older adults remember, or rely on, fewer characteristics to identify the source of items compared to younger adults (e.g., Hashtroudi et al., 1989; Jacoby, 1999; Kensinger, O'Brien, Swanberg, Garoff-Eaton, & Schacter, 2007). Dodson, Bawa, and Slotnick (2007) coined this the *reduced memory hypothesis*. The second theme is that older adults experience false recollections because they have a tendency to incorrectly bind memory characteristics of different events (e.g., Dodson et al., 2007; Ferguson et al., 1992; Henkel et al., 1998). For the purpose of the present research this view will be referred to as the *binding hypothesis*. Although these hypotheses have been contrasted, they are not mutually exclusive, and may both somehow account for age-related source memory impairment (Dodson et al., 2007).

The reduced memory hypothesis. Consistent with the reduced recollection account is the more general finding that older adults are less likely than younger adults to remember details about past events (e.g., Colsher & Wallace, 1991; Craik & Jennings, 1992; Park & Minear, 2004; Salthouse, 2010, as discussed previously). Furthermore, studies have found that older adults are less able than younger adults at remembering memory characteristics that would help attribute the correct source to a memory. For instance, evidence for age-related deficits have been found for memory of sensory information (Kausler & Puckett, 1980, 1981), spatial characteristics (Cherry & Park, 1993; Light & Zelinski, 1983; Park, Puglisi, & Lutz, 1982), temporal features (Kausler & Wiley, 1990; Naveh-Benjamin, 1990), semantic details (Hess, 1984; Rabinowitz, Craik, & Ackerman, 1982), and cognitive processes (Fairfield & Mammarella, 2009).

Hashtroudi, Johnson, and Chrosniak (1990) compared older and younger adults (mean age 19.8 and 68.7, respectively) on their ability to recollect features associated with their memory for common tasks (e.g., packing a picnic basket) that were either perceived (the participant performed the task) or imagined (the participant was read a descriptive transcript of the task). Participants' recollection of memory features was assessed using items from the Memory Characteristics Questionnaire (MCQ; M. K. Johnson, Foley, Suengas, & Raye, 1988), a measure which contains questions regarding recollection of visual, spatial, and temporal features of a memory, and emotions and thoughts that were experienced whilst the event took place. Each item in the MCQ is rated on a 7-point scale to indicate the level of recollection for the specific features. Hashtroudi et al. found that older adults, compared to younger adults, were less likely to remember specific perceptual and contextual details (important for making source decisions) of a memory, and were more likely to remember the thoughts and feelings that had accompanied the memory. Older adults' memory for thoughts and feelings may have made the source of perceived and imagined events seem more similar than if more perceptual and contextual details had been recollected. Similarly, Norman and Schacter (1997) used a DRM false memory task and at recall asked participants questions about different aspects of their memories. They found that younger (mean age 19) adults had better memory for contextual and sound details than older (mean age 67) adults, indicating younger adults were more able to recollect source-specifying cues.

The binding hypothesis. Henkel et al. (1998) believed age-related deficits in source memory were due to older adults having problems accurately binding features into complex memories, and that the bonds between features and memories were not strong. Based on the binding view, Henkel et al. (1998) hypothesised that older adults would have less accurate source memory than younger adults as the similarity between imagined and perceived memories increased. To examine their hypothesis, older (mean age 74.2) and younger (mean age 20.2) participants were presented with an image of a single item, and then saw or were asked to imagine an item that was physically related (e.g., a lollipop and a magnifying glass), conceptually related (e.g., a banana and an apple), or unrelated (control items, e.g., a clothes hanger and a screwdriver) to the initial item. The results showed that older adults' source memory was poorer than younger adults, but recognition memory was comparable across age groups. Both age groups made more source memory errors for related compared to unrelated items. Specifically, older adults were more likely to incorrectly claim conceptually related pairs were perceived compared to unrelated item pairs. Both younger and older adults were more likely to incorrectly claim physically related items were perceived compared to unrelated item pairs. Henkel et al. believed their findings indicated that source memory judgements can be influenced by information from other memories, and that this deficit is greater in older adults. These errors may occur because the features of a memory (e.g., shape) are only loosely bound to the context of their occurrence, and loosely bound features may be incorrectly attributed to another memory, and thus influence source memory judgements. For instance, a lollipop being incorrectly recognised as perceived instead of imagined, when the item perceived was in fact a magnifying glass (Henkel et al., 1998).

Similar to Henkel et al.'s (1998) research, McDaniel et al. (2008) had older (mean age 76.3) and younger (mean age 19.0) adults perform, imagine, or perform and imagine simple actions (e.g., rolling a toy car across a table). Their results found older adults were more likely than younger adults to remember actions that had only been imagined as actions that had been both imagined and performed, and remember performed and imagined items as only being performed. The authors believed these results indicated that age-related deficits in correctly binding memory features to a memory record, resulted in sensory information from performed actions seeming similar to sensory information from imagined actions.

Ferguson et al. (1992) found results suggesting older adults have difficulty correctly binding memory features, particularly subtle features or complex memories when there are multiple cues to identify the source. Ferguson et al. conducted three experiments in which the sources of spoken words were manipulated. In their first experiment participants were presented words from either two female speakers, or a female and a male speaker. In the condition with two female speakers older adults' (mean age 70.1) source memory was poorer than younger adults' (mean age 20.0), whilst in the condition with a female and male speaker no age differences in source memory were found (recognition memory was equivalent across age groups and conditions). These results indicated older adults were unable to effectively use subtle source cues (two female speakers), but were able to effectively use distinctive source cues (a male and female speaker).

In Ferguson et al.'s (1992) second and third experiments their goal was to examine if performance would change if a spatial cue was added. One of these experiments employed a female and male source, the other employed two female sources. Both experiments included a condition with a spatial cue by having the sources sit in opposite corners of the room with a different background (a large potted plant or a colourful print), compared to the condition without a spatial cue, in which the sources sat next to one another. The results demonstrated that when there was no spatial cue and sources were different in gender, older and younger adults' source memory was comparable, but when a spatial cue was added source memory improved only for younger adults (mean age of the groups were 69.8 and 20.0 years). When the gender of the two sources was the same, both age groups' source accuracy improved equally when a spatial cue was available (in this experiment mean groups ages were 18.9 and 70.2 years). These results led Ferguson et al. to conclude that younger adults can coordinate multiple cues to attend or select the most effective cue, or combination of cues, to improve performance. By contrast, older adults are less effective at using multiple source cues, possibly due to deficits in correctly binding memory features, and any gain made from additional contextual information is at the expense of an alternative contextual detail, resulting in difficulties recollecting source.

Dodson et al. (2007) believe older adults have poorer source memory than younger adults because older adults have a tendency towards experiencing convincing misrecollections. Misrecollections are defined as the subjective experience of incorrectly remembering the source of an event; that is, remembering one source presented an item when in fact it was another source. They are convincing because participants confidently believe they are remembering the correct source when they are not. Dodson et al. outline two potential mechanisms that contribute to the occurrence of misrecollections and thus source memory errors. The first mechanism holds that binding errors occur during encoding as features of one memory become incorrectly bound to another memory. The second mechanism holds that during retrieval, features of a target memory are activated as well as features of similar non-target memories. Features of non-target memories may then be confused for features of the target memory.

Source decision/judgement processes. In addition to the binding and the reduced memory hypotheses, researchers have examined the possibility that age-related differences in source memory abilities are due to the decision process that is employed when making a source judgement. As described above, the decision process requires source-monitoring judgements to be made based on the differences between memory characteristics of sources and/or the match between activated schemas corresponding to particular sources and memory characteristics (M. K. Johnson et al., 1993). It is thought that, generally, source-monitoring decisions are quick and made automatically based on the characteristics of activated memories, and sources are retrieved during recall without individuals being aware of the decision process (M. K. Johnson et al., 1993). Individuals also have an acceptance criterion; that is, the information they require before assigning a specific source to a memory (Lindsay & Johnson, 1989). An acceptance criterion is multifaceted and can differ in the amount and type of information needed to make a source judgement. For example, a reasonably liberal standard could rely purely on a sense of familiarity. A stringent standard could require the recollection of specific perceptual details of the source (M. K. Johnson et al., 1993).

Multhaup (1995) conducted research using the false-fame paradigm to investigate the influence an acceptance criterion has on age-related source memory performance. The false-fame paradigm was developed by Jacoby, Kelley, Brown, and Jasechko (1989). In the paradigm participants are exposed to a list of non-famous names and told the names are non-famous. Later participants are asked to rate a second list of names as to whether or not each name is famous. The second list contains non-famous names participants were and were not exposed to earlier and famous names. Multhaup conducted two experiments using false-fame paradigms; the first used the standard false-fame paradigm just described, which is thought to encourage a liberal acceptance criterion when judging source. The second experiment, believed to foster a reasonably stringent acceptance criterion, used the standard false-fame paradigm, and included a source decision task. Participants were asked to categorise the names into famous, non-famous presented in the earlier list, or non-famous not presented in the earlier list. In the first experiment older adults (mean age 70.3) made more false fame errors, judging previously presented non-famous names as famous, than the younger adults (mean age 19.2). In the second experiment, the older participants performed as well as the younger adults, and both age groups' source accuracy benefitted from them being encouraged to use more stringent acceptance criteria. Based on these results, Multhaup asserted that when making source judgements, encouraging stringent acceptance criteria will reduce source misattributions and eliminate age differences. Unfortunately, little research exists on age-related differences in the decision processes used for source-monitoring. Nonetheless, this topic will be returned to in relation to source-monitoring in false memory tasks, but first is an introduction to memory mistakes and false memories, and a discussion of their age-related differences.

Memory Distortions and False Memories

Based on evidence presented in the previous section it appears that there are age-related source memory deficits independent of other memory problems. The source deficits seen in older adults likely decrease their performance in most false memory tasks, not only the DRM paradigm. The following section provides background information about false memories, and outlines research that has examined different types of false memory paradigms, with particular attention given to the DRM paradigm, and age-related differences.

Background. Curiosity about and attempts to understand mistakes in remembering can be seen as early as in the musings of Aristotle with his four laws of association: (a) the law of contiguity (items or events that occur temporally or spatially close will tend to be remembered together), (b) the law of frequency (the more often items or events are linked the greater their association will be), (c) the law of similarity (if item or events are similar the thought of one will lead to thinking of the other), and (d) the law of contrast (experiencing or recollecting an item or event may lead to recollecting an opposite item or event; Sorabji, 1972). Aristotle's ideas had no experimental support until Ebbinghaus (1885/1964) introduced a method that allowed the accuracy of memory to be examined. Ebbinghaus learnt numerous non-word trigram letters (e.g., DAX, BOJ, and YAT), which are known as nonsense syllables. After learning the trigrams he later tested himself for his retention. By using unfamiliar learning materials, materials that would not have been experienced outside of the experiment, Ebbinghaus was able to control the information encoded during learning, and thus determine if an item was correctly retrieved at test. In effect, Ebbinghaus pioneered the experimental method used in psychology (Intons-Peterson & Best, 1998; Roediger & McDermott, 2000a). Prior to Ebbinghaus' research, discussion of memory focused on introspective recollections of previous experiences. The problem with this approach was that, usually, there was no reliable way to determine the accuracy of the recollections (Schacter, 1995).

Once the means to objectively examine memory were established, research into errors of omission began and continues to this day. The earliest research that demonstrated evidence of memory distortions was conducted during the 1900s and focused on examining children's eyewitness testimony. For example, French psychologist Alfred Binet (1900) showed children various items and then tested their memory for the items with and without the presence of misleading questions. Binet found that when children were asked misleading questions they produced many errors as they complied with the researchers suggestions. When the children were asked neutral questions few errors occurred. Additionally, Münsterberg's (1908) publication, *On the Witness Stand,* provided a number of examples emphasising the inconsistent nature of eyewitness reports. This early research into eyewitness memory has been confirmed through further research (Loftus, 1979; Wells, 1978), and it continues to be a popular topic of study (i.e., Dodson & Krueger, 2006; Wade, Green, & Nash, 2010; Wise, Pawlenko, Safer, & Meyer, 2009).

Although the early 1900s provided observations and suggestions regarding memory distortions, it was not until the 1920s and 1930s that experimental evidence into the nature and origin of memory distortions started to appear. Many authors (e.g., Deese, 1959; Gallo et al., 1997; Norman & Schacter, 1997; Roediger & McDermott, 1995) credit Sir Frederic Bartlett with being the first scholar to experimentally investigate false memories, and the publication of Bartlett's (1932) classic monograph, *Remembering,* is claimed to be the most important development of that era (Schacter, 1995). In his memory research Bartlett had participants learn stories; the most well-known being a Native American folktale entitled 'The War of the Ghosts'. After learning the story participants were asked to recall the story several times after different intervals. Bartlett's results showed distortions in participants' memories across the repeated recall attempts. Participants often remembered events that were general and made sense, or events that would be expected in such a story, but were not part of the initial story. Bartlett concluded that recollection is an active and reconstructive process that is driven by one's schemas (abstract and organised mental representations of prior knowledge that underpin one's understanding of the world). Errors can, and do, occur because overall themes are easily remembered, but not specific details. Then, during recollection, missing details in a memory are filled by the individual's schemata and the general theme of the memory. Bartlett asserted that schemas influence what is taken from experiences and how they are remembered. Therefore, the act of remembering is fundamentally social, and memory is inescapably altered by the attitudes and needs one holds.

Bartlett's (1932) work did not have much influence or gain much interest at the time it was conducted and published – it was during the 1970s, after the publication of Neisser's (1967) *Cognitive Psychology* and the beginning of the cognitive revolution that research inspired by Bartlett began to occur (Roediger & McDermott, 2000a). Nonetheless, during the decades after Bartlett's work, occasionally research which analysed memory errors was conducted, but, rather than examine the unreliability of memory, the research generally investigated a theoretical idea. The error aspect was a secondary focus, and only included to examine another feature of memory. For instance, Deese (1959) studied associative processes by examining recall errors. His aim was to predict the occurrence of intrusion errors; words not presented during study that individuals recalled at test. Deese tested memory for word lists using a free-recall test after a single-learning trial. His investigation found that there were some lists of words that reliably produced false recall of a specific word, which were termed *critical words*. Like Bartlett's work, initially the importance of Deese's research went overlooked, until 1995 when Roediger and McDermott replicated, extended, and confirmed Deese's research. Roediger and McDermott replicated Deese's findings using recall and recognition tasks, they expanded Deese's word lists from 12 to 15-items, and developed 24 15-item lists. The research showed participants produced high levels of false recall and recognition of the critical words. In fact, critical words were falsely remembered at almost the same rate as studied words. From this study the DRM paradigm evolved to become one of the most commonly used techniques to elicit false memories.

Age and false memories. From the discussions above, and those that follow concerning age-deficits in the memory processes that underpin false memories, it would seem reasonable to assume that older individuals are more susceptible to false memories than their younger counterparts. However, the findings are not clear cut. Empirical research suggests that older adults are sometimes, but not always, more likely than younger adults to recall or recognise events that did not occur. Some of the early research into false memories and ageing was conducted using the false-fame paradigm (Jacoby et al., 1989) and illustrated that older adults showed a pronounced false-fame effect, greater than that of younger adults (Dywan & Jacoby, 1990; Jennings & Jacoby, 1993). Additionally, using the eyewitness suggestibility paradigm[2] G. Cohen and Faulkner (1989) found that in the group of participants who were misled the older adults made significantly more recognition errors, and had greater confidence in their incorrect responses for misleading information than the younger group. For the control groups, there were no age group differences in recognition responses. Compared to the control group, misled older adults made 29% more errors, while misled younger adults made only 15% more errors. Other researchers have also found that

[2] In a classic eyewitness task (e.g., Loftus, Miller, & Burns, 1978), participants view a film or a slide sequence of a forensically relevant event (e.g., a car crash). Afterwards, half of the participants (experimental group) are exposed to false information regarding the event, in an attempt to mislead them. The other participants (the control group) are exposed to only true information about the event. Generally, results indicate a significant proportion of the participants in the experimental group incorrectly recall or recognise the false information as being part of the original event (Loftus, 1979).

older adults are disproportionately influenced by exposure to misleading information (e.g., Dodson & Krueger, 2006; Loftus, Levidow, & Duensing, 1992; Mitchell et al., 2003; Mueller-Johnson & Ceci, 2004)

Age and DRM false memories. The following section provides an overview of research that has used standard DRM conditions. In the standard DRM recall condition participants are presented (visually or aurally) with one or more DRM lists, then without delay are asked to list all the words they remember. In the standard DRM recognition paradigm, a final recognition test is administered after one or several DRM lists have been presented (visually or aurally) to participants. Although this section focuses on standard conditions, DRM lists can be presented and tested a number of different ways and still cause high rates of false remembering. For instance, generally, a presented DRM list consists of 12-15 associated words, but Gallo and Roediger (2003) presented participants with 5-, 10-, or 15-item DRM lists and found false recognition rates ranged from .55 (5-items) to .68 (15-items). Similar results have been demonstrated using 7-item DRM lists (Sugrue & Hayne, 2006), and 3-item DRM lists (Hancock, Hicks, Marsh, & Ritschel, 2003). The false memory phenomenon is so robust that false memories occur when participants are asked to solve anagrams of DRM list items (Hicks & Marsh, 1999), when DRM items are presented within sentences (Thomas & Sommers, 2005), when DRM words are presented alongside a simple black and white image of the item (Schacter, Israel, & Racine, 1999), when participants are given a thorough warning about the false memory phenomenon (Gallo et al., 1997; McCabe & Smith, 2002; Neuschatz, Benoit, & Payne, 2003), and when testing is immediate or delayed (McDermott, 1996).

Recall. To briefly review age-related findings using the standard DRM recall condition, Gallo (2006) selected data from 18 experiments from 12 journal articles[3] that compared older (mean age 73 years) and younger adults (mean age 21 years). Across the 18 experiments the mean false recall rate of critical lures was significantly greater in older adults compared to their younger counterparts (.40 and .33, respectively), and the mean correct recall rate of studied items was significantly lower in older adults compared to

[3] Balota et al. (1999); Butler, McDaniel, Dornburg, Price, and Roediger (2004); Intons-Peterson, Rocchi, West, McLellan, and Hackney (1999); Kensinger and Schacter (1999); Lövdén (2003); Norman and Schacter (1997); Rybash and Hrubi-Bopp (2000); Thomas and Sommers (2005); Tun, Wingfield, Rosen, and Blanchard (1998); Waldie and See (2003); Watson, Balota, and Sergent-Marshall (2001); Watson, McDermott, and Balota (2004).

younger adults (.47 and .62, respectively). Although when consolidated these 18 experiments demonstrated older adults performed more poorly than younger adults, when the experiments are examined separately approximately one third illustrated no, or very small, age differences of false recall in the basic DRM paradigm (Intons-Peterson et al., 1999; Kensinger & Schacter, 1999; Rybash & Hrubi-Bopp, 2000; Thomas & Sommers, 2005; Tun et al., 1998). Increases in false recall with age are widely reported, but the pattern across individual studies is inconsistent, possibly because it is a small (or noisy) effect and consequently difficult to detect.

Recognition. As with recall, Gallo (2006) selected data from 15 experiments from 10 research articles[4] comparing younger (mean age 20 years) and older adults (mean age 72 years) on false recognition. The selected data were from standard DRM recognition conditions in which recognition tests were not confounded by former recall tests. Correct recognition was significantly lower for older adults compared to younger adults (.73 and .77, respectively). However, false recognition was not significantly different for the younger and older adults (.68 and .71, respectively). In fact, more than half of the experiments found no significant difference in false recognition across age groups (Benjamin, 2001; Budson et al., 2000; Gallo & Roediger, 2003; Intons-Peterson et al., 1999; Kensinger & Schacter, 1999; McCabe & Smith, 2002).

For research in which recall tests are performed ahead of recognition tests (e.g., Balota et al., 1999; Intons-Peterson et al., 1999; Norman & Schacter, 1997; Tun et al., 1998; Waldie & See, 2003) a significant difference *is* found in false recognition between younger (mean age 20 years) and older (mean age 72 years) adults (.71 and .83, respectively). Gallo (2006) argued that this is possibly due to carryover effects, in that the initial recall test provides an additional opportunity to encode or practice for later retrieval. Any true or false items initially recalled would be primed to be subsequently recognised. Research that has directly examined the effects of conducting a recall test prior to a recognition test, has found that false recognition increases by a small amount (approximately 4 to 10%; Roediger & McDermott, 1995; Roediger, McDermott, Pisoni, & Gallo, 2004), or not at all (Payne, Elie,

[4] Benjamin (2001); Budson, Daffner, Desikan, and Schacter (2000); Budson, Sullivan, Daffner, and Schacter (2003); Gallo, Bell, Beier, and Schacter (2006); Gallo and Roediger (2003); Intons-Peterson et al. (1999); Kensinger and Schacter (1999); McCabe and Smith (2002); Schacter, Israel, and Racine (1999); Thomas and Sommers (2005).

Blackwell, & Neuschatz, 1996; Schacter, Verfaellie, & Pradere, 1996). Like recall, it is possible that with false recognition there are concerns with statistical power; that is, the effect is small, and as a result, studies require large numbers of participants to detect significant age-related differences.

In sum, it appears that when researchers utilize DRM procedures, age-related differences are found in false recall more often than in false recognition. This is consistent with research into true memory, as recognition tests provide better retrieval cues than recall tests, but is somewhat unexpected for false memory. Older adults are expected to perform more poorly on false recognition tasks compared to younger adults, because the tasks are affected by errors based on familiarity, which are believed to be greater in older adults than younger adults (Gallo, 2006; Jacoby, 1999; Tun et al., 1998). A possible explanation is that the free recall tests in the DRM paradigm profit more from source-monitoring than the recognition task, because free recall tests are often done immediately after a single lists presentation. By contrast, recognition tests are generally completed after the presentation of multiple lists (a much longer study phase). Age-related source-monitoring impairments would be more obvious in recall than recognition tests, because recall tests require conscious generation and regulation of an individual's response immediately after the presentation of a list. For a recognition task there may be limited information available for successful source-monitoring after the longer study phase (Gallo, 2006).

Theories of False Memories

The previous section has outlined a number of studies that focus on DRM tasks and age-related effects (more of which will be discussed below). The following sections describe and discuss the theories of the underlying mechanisms or processes thought to create false memories, and how the mechanisms can explain false memory effects related to age. The theories that are discussed constitute the two most commonly employed, and opposing, theories for explaining false memories: the activation-monitoring theory (semantic activation and source-monitoring) and the fuzzy-trace theory (gist and verbatim memory traces).

The Activation-Monitoring Theory

The activation-monitoring theory is the most widely endorsed explanation for false memories (Dehon & Brédart, 2004; Roediger & McDermott, 1995; Roediger, Watson, et al., 2001; Skinner & Fernandes, 2009). The theory consists of two mechanisms: semantic activation and source-monitoring. Put simply, the presentation of a DRM list (or other priming task) activates an individual's lexical or semantic system, and as a result related non-presented concepts (such as critical lures) are (unconsciously or consciously) evoked (Roediger, Watson, et al., 2001). During retrieval the individual uses the extensive amount of accessible information to judge where a memory originated from (M. K. Johnson et al., 1993; Roediger, Watson, et al., 2001).

The activation mechanism. The activation mechanism is founded on the implicit associative response model put forward by Underwood (1965). Underwood found that when participants studied one word (e.g., table) false recognition of an associated word (e.g., chair) increased. The interpretation of this finding was that presentation of one word aroused an implicit associative response to another word during encoding. This process is considered implicit because it is thought individuals are not aware that activation of the critical lure is occurring during study. Subsequently, the activated concepts may be stored during study and then be accessible during retrieval (Roediger & McDermott, 1995). In the DRM task, presenting a list of semantically-associated words activates one's semantic system (Roediger, Watson, et al., 2001). This activation spreads to generate an implicit associative response, which for DRM lists is the activation of a word, or words, strongly associated with the list items. It is the evocation of non-presented concepts that are associated with presented items that creates false memories for critical lures (Roediger, Watson, et al., 2001), as the activation from multiple words in the DRM list converge on and prime the non-presented critical lure (Roediger, Balota, & Watson, 2001).

According to Roediger, Balota, et al. (2001) the strongest evidence that the activation mechanism underpins DRM paradigm false memories is that, as the amount of activation for a lure increases (e.g., increasing the number of semantically-associated words presented increases activation), so does recall or recognition of that lure. Robinson and Roediger (1997) presented participants with DRM lists of differing lengths (3, 6, 9, 12, or 15 words) and measured both false and correct recall. The results demonstrated that as the length of the study list increased, the proportion of presented items recalled decreased, and the proportion of critical lures recalled increased. Similar results have been found for false recognition with lists of 3, 5, 9, and 12 words (E. Marsh & Bower, 2004). To examine the results when the number of related items presented was not confounded with list length, Robinson and Roediger completed a second experiment in which filler words, unrelated to the DRM list, were added to bring all of the DRM lists presented to 15 items. Compared to the results of the first experiment, adding the filler items lowered the level of correct recall, but had no effect on false recall. These results suggest that false recall is determined by the total amount of associative strength from a list to the lure item, and not the mean strength, because adding filler words does not change the total associative strength, but decreases the average strength.

Roediger, Watson, et al. (2001) conducted a regression analysis that strongly implicated an activation mechanism. The multiple regression analysis examined eight factors across 55 DRM lists that were related to both characteristics of the DRM lists and the critical lures. The eight variables examined were (a) veridical recall; (b) false recall; (c) inter-item associative strength of the presented words; (d) forward associative strength, the probability a critical lure elicited each presented word in its respective DRM list; (e) backward associative strength (BAS), an index of the mean strength of associated connections from each DRM list word to the critical lure. BAS is obtained by averaging the joint probabilities of each study word eliciting the critical lure in a free association test (Deese, 1959); (f) concreteness of each critical lure; (g) log frequency, the number of times a critical lure is found in print per a million words, transformed to correct for skewness in the frequency distribution; and (h) word length, the number of letters in the critical lure (Roediger, Watson, et al., 2001).

The multiple regression analysis found a high correlation, .73, between BAS and false recall. Of the eight factors examined, BAS was found to be the main contributor to false recall. The researchers interpret these findings as being consistent with the spreading activation model, because the stronger the association between list items and the critical lure the more likely that the critical lure will be activated, and the more likely that participants will falsely recall or recognise the lure (Roediger, Watson, et al., 2001). McEvoy, Nelson, and Komatsu (1999) also found evidence suggesting that BAS plays a role in false memories, and Deese (1959) reported a strong correlation, .87, between BAS and false recall.

A further point to consider is whether or not semantic activation of critical lures is conscious (the lure comes into one's awareness during presentation) or unconscious (the lure does not come into one's awareness). Many scholars believe that both conscious and unconscious factors play a role in activation of critical lures (McDermott, 1997; Roediger, Watson, et al., 2001; Seamon, Luo, & Gallo, 1998), and empirical evidence suggests that both occur. Roediger, Balota, et al. (2001) report that if conscious activation of list items is necessary to create false memories, then presentation rates too fast to allow conscious processing of the DRM items should eliminate the false memory phenomena. To address this idea, Robinson (1998) presented 15-item DRM lists at fast durations: 20, 80, 160, or 320 ms per word. The results demonstrated that even at the fastest rate, when a list was presented within a second, false recall occurred, indicating that unconscious factors play a role in activation. Furthermore, veridical recall and false recall increased in a one-to-one manner as presentation rate slowed. For the four durations, veridical recall was .10, .22, .28, and .31, and false recall was .10, .25, .31, and .33, respectively. Assuming activation increased as the presentation rated slowed, then false memories increased in proportion to activation.

Gallo and Roediger (2002) determined that with very slow presentation rates, false recall is negatively associated with veridical recall. Gallo and Roediger slowed the presentation rate of DRM lists to 500, 1000, and 3000 ms per item. The results showed that correct recall rates increased (.58, .65 and .73), whereas false recall rates decreased (.48, .41, and .28) as the presentation rate slowed. McDermott and Watson (2001) examined fast and slow presentation rates (not either/or as in the research above), employing presentation durations of 20, 250, 1000, 3000, and 5000 ms per DRM item. Their data showed that correct recall increased (.17, .31, .42, .50, and .51) as presentation duration increased. False recall of critical lures followed an inverted U pattern – critical lure recall rose and then fell as the presentation speed slowed (.14, .31, .22, .14, and .14).

The above results indicated that the critical lure can be automatically activated, as false recall occurred at very fast rates in which conscious activation of the lures was unlikely. However, the fact that false memories decreased with longer presentation durations is incompatible with a straightforward activation explanation, which would assume increased presentation durations leads to increases in activation, suggesting other processes are involved. It is possible that with the longer duration of presentation rates recollection of specific information about items (e.g., source, context, or perceptual information) is strong enough that participants can rely on the recollection of such information, and not rely only on activation to make memory judgements (Roediger, Balota, et al., 2001).

Further evidence for conscious activation of critical lures is seen in research that has employed Tulving's remember/know paradigm (1985). In the remember/know paradigm, after participants decide a test item is old, they judge if they can 'remember' the item (recollect contextual aspects of the study phase), or 'know' the item was present (they cannot retrieve contextual information, but the item is adequately familiar for an old response). Applying the remember/know paradigm to DRM tasks has shown that participants often demonstrate high levels of remembering for critical lures they have recalled or recognised (Gardiner, Ramponi, & Richardson-Klavehn, 2002; Roediger & McDermott, 1995). The high levels of remembering for the critical items suggest that participants have become aware of the items during study (Roediger & McDermott, 2000b).

Ageing and the activation mechanism. Few researchers have focused on the relationship between the activation mechanism and age-related differences in false memories. Researchers who do consider the activation mechanism employ Hasher and Zacks' (1988) inhibition hypothesis, described earlier. To recap, the hypothesis proposes that older adults have problems with memory tasks because they cannot remove or restrain irrelevant information from their working memory as successfully as younger adults. Due to an inhibitory deficit, in false memory tasks older adults may experience higher levels of activated relevant and irrelevant information compared to younger adults (Hasher & Zacks, 1988; Tun et al., 1998). However, Dehon and Brédart (2004), Balota et al. (1999), and Tun et al. (1998) have all reported results that indicated lures were activated as often in younger adults as they were in older adults. These researchers argued that the extent of spreading activation in an individual's semantic system remains relatively stable through ageing; therefore, age-related increases in false memories are due to deficits in the monitoring process. For instance, Dehon and Brédart (2004) showed that after hearing DRM lists older adults more often indicated they had heard the lures, compared to younger adults who indicated that the lure had come to mind, but they did not remember hearing it. These results imply that the critical lures were activated during encoding similarly in older and younger adults, but older adults were less successful at source-monitoring compared to younger adults. It is clear from the previous discussion that encoding features play a significant role in the creation of false memories. Retrieval factors, specifically source-monitoring, are important as well.

The source-monitoring mechanism. As defined earlier, source memory is a memory or belief for the origin of information that is encoded in one's memory (M. K. Johnson et al., 1993). In false memory literature source memory errors are also identified as source-monitoring errors (Gallo et al., 1997; M. K. Johnson et al., 1993; Norman & Schacter, 1997; Roediger & McDermott, 1995; Roediger, Watson, et al., 2001; Skinner & Fernandes, 2009). Within the DRM paradigm it is believed that critical lures are activated strongly during encoding; then during retrieval participants retrieve the lures as a presented list item because the lure has been stored with experiential characteristics that are similar to list items (Roediger, Watson, et al., 2001). It is the degree to which characteristics of a critical lure are similar to characteristics of real list items that increases false recognition or recall. For instance, if a lure were to enter awareness during encoding, then was rehearsed along with the presented list words, the associated memory characteristics of the lure would be very similar to presented items (Roediger, Watson, et al., 2001). Roediger and McDermott (2000b) believe evidence for this idea is observed when participants falsely remember the experience of hearing the critical lure (as seen in work by Gallo et al., 1997; Read, 1996; Roediger & McDermott, 1995).

One way to investigate source-monitoring has been to present items via varied sources and employ a source memory judgement within a recognition or recall test. The belief is that source decisions require individuals to more carefully inspect their memory for an item. Consequently, false memory rates should be lower compared to false memory rates in a simple yes/no or old/new recognition test (Hicks & Marsh, 1999, 2001; Multhaup & Conner, 2002). Surprisingly, such research into the DRM phenomena suggests that encouraging participants to use a source-monitoring strategy does not necessarily lead to reductions in false recognition. Hicks and Marsh (1999) conducted five experiments and compared DRM false *recall* when the number of sources from which items were learnt was one or two, and when source judgements were or were not required. Three different source discriminations were employed by Hicks and Marsh: (a) internal-internal (participants were asked to judge the pleasantness of half of the presented DRM list items and rated the other half for how frequently each word had been encountered in the last few weeks); (b) internal-external (participants solved anagrams for half of the DRM words and were verbally or visually presented with the other half); and (c) external-external (half of the words were presented by a female speaker, the other half by a male speaker). Their research showed that false recall decreased only when source discriminations were internal-external. There were no differences in false recall when comparing the condition with only one source to either the internal-internal or external-external source discrimination conditions. Furthermore, when participants were given the option to respond that they were not sure of the source for an item, they continued to incorrectly judge one of the sources to be the source of the falsely recalled items.

Hicks and Marsh (2001) examined the influence of internal-external and external-external source discriminations on DRM false *recognition* in a standard old/new recognition task and a source task. In Experiment 1, DRM items were either heard or generated through anagrams. In Experiment 2, items were either heard or seen. In their third experiment, items were presented by a female or male. After being presented with the six DRM lists, half of which were presented by each source, participants completed a standard old/new recognition test only, or a standard old/new recognition test with an added task in which participants were to judge the source of items they perceived as old. The results demonstrated that across all three experiments false recognition of critical lures was higher when participants were asked to judge the source of items, compared to when participants completed the standard recognition task.

Other research has found *reductions* in false memories when participants are requested to judge the source of items. Multhaup and Conner (2002) compared a group who were warned about the DRM phenomena before study and completed a source-monitoring test, to a group who were not warned but completed the source-monitoring test, and a group who were not warned and did a standard old/new recognition task. Multhaup and Conner suspected that prior research into the DRM paradigm that included a source task did not find declines in false memories because they had not included the correct source. Therefore, participants’ source options at test were: I did not hear this word, I did not hear this word but generated it on my own, I heard this word and generated it on my own, and I heard this word. Multhaup and Conner expected that including the correct source (i.e., I generated it on my own) would make participants more closely examine their memories for items, and would make participants use a strict acceptance criterion. Over their experiments the groups that did the source task had much lower false recognition rates than the group who took part in the standard condition. Because Multhaup and Conner had participants do individual study-test trials with 20 different DRM lists, instead of having participants complete one final recognition test after the presentation of several lists, it is unclear if the reductions were due to source-monitoring at test, or an identification strategy employed during study after subsequent study-test trials, or both. Research similar to Multhaup and Conner's using the DRM paradigm has yet to be replicated to clarify the cause of the false recognition reductions. Similar research has been conducted using the eyewitness suggestibility paradigm with younger adults (Lindsay & Johnson, 1989; Zaragoza & Lane, 1994), and comparing younger and older adults (Multhaup, De Leonardis, & Johnson, 1999). Both of these studies showed that including the correct response in a source task improves older and younger adults' performance.

The literature describes two possible reasons why source-monitoring manipulations do not always reduce false memories. The first explanation is that participants, in any type of condition, may spontaneously use a monitoring process (this may not specifically be source-monitoring) without special instructions, or without the requirement of source judgements (Gallo, 2006). If participants already employ a monitoring strategy, instructing them to do so, or instructing them to focus on specific characteristics of presented items, may not provide them with additional information that enhances their monitoring process. Lampinen, Meier, Arnal, and Leding (2005) provided evidence suggesting participants do utilise monitoring strategies without being instructed to. Their procedure employed a standard DRM task, except that participants were required to think aloud throughout the experiment. Evidence of the use of a recollection-based monitoring process was found for a number of the test trials in which participants correctly decided the critical lure was not studied.

Another reason why source-monitoring manipulations may not be beneficial is because between the items there are limited item-specific details to recollect (e.g., how items were presented), and limited semantic variation, making it difficult to formulate accurate judgements (Gallo, 2006). For example, in research that employs standard DRM procedures all DRM lists are presented by means of the same source (e.g., a male voice), and have similar semantic qualities. Both of these aspects influence false memories in participants of any age because there are fewer item-specific qualities to discriminate lures from studied items (Gallo, 2006). Results consistent with the idea that source-monitoring manipulations are not beneficial because limited item-specific differences exists among items, was shown by Hashtroudi et al. (1989). In Hashtroudi et al.'s research both older and younger adults performed equally discriminating the source of words (not DRM words) when they originated from either an internal (participants said a word, thought of themselves saying a word, or generated the word from an anagram) or external source (listening to someone else say a word), which provided a high-level of contextual variation between items. Similar results have been found by Ferguson et al. (1992).

Additionally, scholars believe that providing a warning to participants about the creation of false memories reduces their occurrence, as individuals can use a monitoring process to identify activated lures as non-studied words, and then reject them (Gallo et al., 1997; Neuschatz et al., 2003). Gallo et al. (1997) employed a condition in which participants were told detailed information about the DRM lists (i.e., how they were composed and that previous research has revealed they lead to high false memory rates), and the false recognition effect (i.e., the nature of the effect, an example of a DRM list and its corresponding lure, and that the aim in the research was to minimise false recognition). During the warning, participants also heard a DRM list, completed a recognition test, and had the critical lure identified for them. Neuschatz et al. (2003) conducted DRM research using a condition in which participants were informed that each list contained related words, associated to a common word that tied all of the words in the list together, that this word may or may not have been presented during study, and they had to attempt to identify the common word and whether or not it was presented during study. The researchers took the participants through an example as well, identifying the critical lure after the test and stating that often the lure is incorrectly remembered as being presented. Both of the studies (Gallo et al., 1997; Neuschatz et al., 2003) described above found that those who were warned produced fewer false memories than those who were not, but false memories were not completely eliminated in the warned participants. False memories reduced from .81 to .46 in Gallo et al.'s research, and from .49 to .30 in Neuschatz et al.'s study. Similar results have been reported by McDermott and Roediger (1998).

Unlike the studies described above, which show younger adults consistently decrease false memories when provided with a warning, findings have been inconsistent when investigating the impact of providing older adults with a warning before the study phase of a DRM task. McCabe and Smith (2002) and Watson et al. (2004) found older adults benefited from warnings, though not as much as younger adults. Dehon and Brédart (2004) found older adults did not benefit from warnings. Based on these results it appears that people cannot always bring false remembering processes under successful conscious control (McDermott & Roediger, 1998), but warnings can improve performance in both younger, and sometimes older, adults (Gallo et al., 1997; McCabe & Smith, 2002; Watson et al., 2004).

Ageing and the source-monitoring mechanism. Based on earlier discussions, it seems apparent that the source memory/monitoring abilities of older adults are not as effective as those of younger adults (G. Cohen & Faulkner, 1989; Ferguson et al., 1992; McIntyre & Craik, 1987; Schacter et al., 1991; to name a few). One reason believed to explain the age-related deficit is that source-monitoring involves the use of controlled processes (Benjamin, 2001; Skinner & Fernandes, 2009), and evidence suggests older adults' ability to use controlled processes are often impaired compared to younger adults, whereas automatic influences in memory remain stable across the lifespan (Jacoby, 1999; Jacoby, Jennings, & Hay, 1996; Jennings & Jacoby, 1993, 1997; Shiffrin & Schneider, 1977). This pattern of impairment and stability results in higher rates of false memories in older compared to younger adults (Balota et al., 1999; Norman & Schacter, 1997; Skinner & Fernandes, 2009).

Skinner and Fernandes (2009) investigated the influence of controlled and automatic proccesses on age-related DRM performance. Older (mean age 74.0) and younger (mean age 19.1) adults were presented with DRM lists one or three times. The results showed that false recognition in the older adults was higher while younger adults' false recognition was lower when DRM list presentation was repeated. Both age groups had higher correct recognition when the lists were presented three times compared to once. These results suggest that the repeated exposure to the DRM lists increased the automatic influence of familiarity and/or activation of the critical lure for both age groups, but the older adults were unable to counter these automatic influences by using controlled recollection processes such as source-monitoring. Similar results have been found by Kensinger and Schacter (1999) and Budson et al. (2000), discussed below, as these researchers believe their findings support the fuzzy-trace theory over the activation-monitoring theory.

Researchers have claimed that older adults lack the resources necessary to (a) separate true memories from false memories by considering the differences in the memory characteristics of the two (Norman & Schacter, 1997; Thomas & Sommers, 2005), and (b) mark critical lures as self-generated during encoding (Skinner & Fernandes, 2009). For instance, Norman and Schacter (1997) demonstrated that older adults (mean age 67.0) are less able to use perceptual and contextual features (measured using the MCQ) to discriminate between presented items and lures, compared to younger adults (mean age 19.0). From the above section it is clear that the activation-monitoring theory is a well-researched theory of false memories, though as stated earlier, activation-monitoring is not the only theory commonly employed to explain false memories. The second theory is the fuzzy-trace theory.

The Fuzzy-Trace Theory

The fuzzy-trace theory is a dual-process model that originated as a framework to explain reasoning and decision making, and has since been applied to a number of memory phenomena (Brainerd & Reyna, 2002a, 2005; Reyna & Brainerd, 1995a). The theory has a number of important properties, but the main explanation for false memories is that there are mechanisms in memory that function in opposition to one another (verbatim and gist memory traces). This idea differs from dual-process theories of true memory where recollection and familiarity support each other to strengthen correct recognition, and correct recall is supported by both direct access and reconstruction (Brainerd & Reyna, 2005). Verbatim memory traces are integrated representations of an item's surface content and item-specific information that is associated with subjective recollection. Gist memory traces are representations of overall similarity, meaning, and relationships between presented items. According to the fuzzy-trace theory there are separate storage and retrieval functions for these two trace systems. During presentation stimuli are encoded using both verbatim and gist traces and the two traces are assumed to be stored in parallel (Brainerd & Reyna, 2002a; Gallo, 2006). In false memory tasks, the simultaneous processing and storage of both gist and verbatim traces means participants retain a considerable amount of information regarding an item's meaning, even when they are unsuccessful at fully processing the item's surface form (Brainerd & Reyna, 2005).

Brainerd and Reyna (2002a, 2005) assert that memory performance is based on the accessibility and retrieval of both verbatim and gist traces, and the representations accessed at test depend on the retrieval cues provided. Retrieval of verbatim traces stimulates vivid recollection of a target's earlier presentation. The inability to retrieve verbatim traces is a result of the disintegration of item-specific features of the memory. Verbatim retrieval is preferred when verbatim traces are strong compared to gist traces, and gist retrieval is preferred when gist traces are strong compared to verbatim traces (Brainerd & Reyna, 2002a). When verbatim traces are accessible, correct items are better cues for verbatim than gist traces, and recall and/or recognition of correct items is predominately due to verbatim retrieval over gist retrieval. However, the accessibility of verbatim traces deteriorates much faster than the accessibility of gist traces. As time continues, there will be a shift from reliance on verbatim traces to a reliance on gist traces (Brainerd & Reyna, 2002a, 2005). Regardless of the accessibility or inaccessibility of verbatim traces, in false memory tasks related lures are superior cues for gist than verbatim traces, and false memories in recall and/or recognition tests are mainly due to the retrieval of gist information (Brainerd & Reyna, 2005). False memories do not involve the encoding of exact content, but rather rely on the encoding and retrieval of semantic features and an overall theme (Brainerd & Reyna, 2002a).

To further elaborate, retrieval of both gist and verbatim traces are responsible for true memories, whereas they have opposing influences on false memories (Brainerd & Reyna, 2002a). Verbatim retrieval accounts for true memories because details of the experience can be explicitly recollected. Gist retrieval accounts for true and false memories in the same way, through familiarity (Brainerd & Reyna, 2002a). Gist retrieval reinforces true memories because the meaning of an item is familiar; even when an item's experiential details cannot be recollected it is familiar enough to be considered a true memory (Brainerd & Reyna, 2002a). A critical lure is consistent with the gist of a DRM list. At test the accessibility and retrieval of gist information results in a signal that the lure had been presented (Gallo, 2006). Conversely, retrieval of verbatim memory traces suppresses recall and/or recognition of false memories by counteracting the perceived familiarity. For example, in a recollection reject strategy, intrusions and false memories can be suppressed if particular incorrect items are edited from memory representations because verbatim traces of the respective correct item are accessed (e.g., specifically remembering that you drank a coke for lunch, not a sprite). Furthermore, if a strict set of rules are used in the decision process, false items will be rejected if their memory representations do not contain clear verbatim details of the experience (e.g., an individual only accepting items that they have a specific aural recollection of its presentation; Brainerd & Reyna, 2002a, 2005).

Ageing and gist and verbatim memories. The basic premise on which the fuzzy-trace theory explains older adults susceptibility to false memories compared to younger adults is that older adults have both impaired recollection of verbatim traces, and preserved retention of gist traces (Budson et al., 2000; Kensinger & Schacter, 1999; Tun et al., 1998). For example, in an attempt to increase recollection of item-specific details (verbatim traces) for presented items Kensinger and Schacter (1999) repeated study-test trials of three 15-item DRM lists five times. Two experiments were conducted, the first examined recall and the second recognition. The results showed that younger adults (mean age 19.9 and 19.3, respectively) were able to decrease false recall and false recognition rates across the study-test trials. The older adults' (mean age 67.4 and 68.2, respectively) false recall and false recognition remained stable across the trials. Both older and younger adults increased their correct recall and recognition across the trials.

Budson et al. (2000) examined age-related differences in DRM performance with five repeated study-test trials. Across the five trials younger adults' (mean age 19.4) false recognition decreased, and older adults' (mean age 74.3) false recognition fluctuated; by the last trial false recognition was just barely significantly less than at the first trial. Both groups' correct recognition increased across the five trials. Kensigner and Schacter (1999) and Budson argued that their results indicated younger adults had the ability to use the repeated trials to improve their use of item-specific information and offset the accumulation of gist memory traces across trials, thus suppressing false memories, but older adults did not have this ability and were more reliant on gist memories. Although, Kensigner and Schacter and Budson focused on the fuzzy-trace theory explaining their results, they did acknowledge that age-related source memory confusions may have played some part in their findings. Similar research and findings were discussed earlier (refer to page 35) in the context of activation-monitoring.

To investigate the role of gist and verbatim memory in the DRM paradigm Tun et al. (1998) conducted research that compared older and younger adults on tasks that emphasised or deemphasised the usefulness of gist-based strategies; that is, the efficiency of making recognition decisions based on thematic associations, gist traces, between DRM items was altered. In their first experiment gist-based processing was advantageous in the recognition test. To the participants it would have seemed that recall and recognition decisions could be made purely based on the semantic relationship between the items. This is because, unless specifically warned, participants are not aware that the research they are taking part in is examining the creation of false memories. Younger (mean age 20.3) and older (mean age 70.1) participants heard a single 12-item DRM list which was followed by an immediate recall test, then an immediate recognition test. The recognition test contained unrelated[5] words, the critical lure, and presented items. This procedure was repeated with 10 different lists. The researchers also measured the response latencies for recognition. The results demonstrated older and younger adults had equal recall for the critical lures, but correct recall was significantly lower for the older than the younger adults. The recognition results showed no significant age differences in true or false memories. Overall, older adults were slower to respond than the younger adults, but both age groups responded equally as quickly to correct items as they did to critical lures. Therefore, when a recognition test can rely on gist-based processes the performance of younger adults (who are thought to normally rely on verbatim- and gist-based processes) resembles that of older adults (Tun et al., 1998).

In the second experiment the researchers altered the recognition test so verbatim and gist processes were required to make efficient recognition decisions. Tun et al. (1998) accomplished this by excluding from study not only the critical lure but also three weak lure words from each DRM list, which were then present in the recognition tests. That is, participants heard a 12-item DRM list, and then completed an immediate recall test followed by an immediate recognition test that contained presented items, the critical lure, three weak lures, and unrelated words. The recognition results showed no age-related differences in correct recognition, but false recognition of critical and weak lures was higher in the older (mean age 72.8) than the younger (mean age 18.8) adults (similar results have been found by Waldie and See, 2003). Additionally, false recognition was equal to correct recognition in the older group, but younger adults recognised more true items than false items. This pattern was matched by the response latencies; older adults responded yes to critical lures and presented items at the same speed, while younger adults responded yes to critical lures more slowly than for presented items. Older adults were also slower to reject weak lures than younger adults. According to Tun et al., the results from Experiment 2 indicate that

[5] Unrelated words (or unrelated items) are test items that were (generally) not presented during study, and were not semantically associated with the DRM word list(s) that were presented.

when the efficacy of gist-based memory processes is decreased, age-related increases in false memories occur, because older adults remain reliant on gist more than younger adults who utilise both gist and verbatim traces.

In the third experiment, Tun et al.'s (1998) aim was to replicate the results found in Experiment 2 when the *study* condition was not conducive to gist memory processes. They presented older and younger (mean age 19.9 and 67.0, respectively) participants with 20 randomly ordered words, from four different DRM lists (each DRM list contributed five words). Presenting lists in a random order means their thematic associations are less apparent and less likely to favour gist processes at encoding than if they were presented in a blocked fashion. The free recall test showed no age-related differences in false recall of critical lures, both age groups' correct recall was higher than false recall, and the younger adults' correct recall was higher than the older adults'. False recognition was equal across age groups. Correct recognition was higher in the younger, compared to older adults. These recognition results suggest that when items are presented randomly, making the associations between items obscure, having to employ effortful processes to form gist representations disadvantages older adults on their correct recognition performance (correct recognition in Experiments 1 and 2 was equal across age groups). That is, the preservation of gist processes negatively affects verbatim processes (Tun et al., 1998). The previous section has provided a thorough discussion regarding the fuzzy-trace theory as an explanation for both false memories and age-related differences in false memories. However, many of the research results discussed in this section can potentially be explained by the activation-monitoring theory, and vice versa. The following section will examine common features of the two theories that make it difficult for research to definitively support one theory over the other.

Common Characteristics within the Two Theories

Although writers often favour either the fuzzy-trace theory or the activation-monitoring theory, both are used to explain a number of false memory phenomena as well as the DRM paradigm. These phenomena include (but are not limited to), false memories using the eyewitness misinformation paradigm (Lindsay, Gonzalez, & Eso, 1995; Reyna & Titcomb, 1997), false memories for prose material (M. K. Johnson, Bransford, & Solomon, 1973; Reyna & Kiernan, 1994), false recovered memories (Brainerd & Reyna, 2002a; Lindsay

& Read, 1994), and false recognition reversal (when unrelated items have a higher false recognition rate than related items not presented at study; Brainerd, Reyna, & Kneer, 1995; Lindsay & Johnson, 2000). In DRM research, authors will quite often examine their results using both of the theories because it is difficult to design experiments that can decisively support one over the other. This is not surprising since there is substantial overlap in the claims made by the two theories.

One claim that both frameworks make is that older adults will perform more poorly on false memory tasks than younger adults. Nonetheless, many researchers have found no age-related differences in false memory performance (e.g., Benjamin, 2001; Budson et al., 2000; Gallo & Roediger, 2003; McCabe & Smith, 2002; Sim, 2010b; Thomas & Sommers, 2005; Tun et al., 1998). Both theories can explain such findings. The activation-monitoring explanation is that the task did not encourage participants in either/any of the age groups to use a source-monitoring strategy to establish why an item was familiar/activated (Benjamin, 2001; Kensinger & Schacter, 1999; Sim, 2010b), or the source-monitoring strategy employed was unable to differentiate sources, or unable to help younger adults reduce false memories (Gallo & Roediger, 2003; Sim, 2010b). For example, Sim (2010b) found high levels of false recognition in young, middle-aged, and older adults (.69, .69., and .77, respectively), suggesting that the DRM lists participants heard strongly activated the corresponding critical lures in the majority of participants, and most of those participants did not, or could not, use a source-monitoring strategy to identify critical lures as non-presented items in the recognition test.

Under the fuzzy-trace theory, when no age-related differences are found it is because accurate performance in the false memory test appeared to be able to be achieved by employing gist strategies (McCabe & Smith, 2002; Tun et al., 1998). This, it is claimed, is because participants often are not aware that the research they are taking part in is examining the creation of false memories. Consequently, participants will employ the most effortless strategy – a gist strategy – which means adults, regardless of age, will perform similarly. For instance, along with high levels of false recognition, Sim (2010b) found a strong positive association between correct recognition and false recognition of critical lures in each age group. According to the fuzzy-trace theory, if participants had been employing verbatim and gist memory traces, such high rates of false memories would not have been expected and there would have been a negative relationship between true and false recognition (Brainerd & Reyna, 2002a). Sim's research suggests individuals in each age group relied heavily on gist memory traces, with little or no consideration given to verbatim memory traces.

Another reason why source-monitoring and fuzzy-trace frameworks are intrinsically linked is because source information – as it is commonly operationalised and manipulated in experiments – represents a verbatim memory trace. As such, it is a surface detail that can be differentiated from gist traces, and like all verbatim traces it becomes fragmented and inaccessible over time. Because verbatim traces become inaccessible more rapidly than gist traces, source information can become disconnected from its original experience and become incorrectly associated to another experience, which is called a source confusion error (Reyna & Lloyd, 1997). Researchers indicate that source details may decay quicker than other verbatim details, because some verbatim details remain intact, whereas a memory's source may be forgotten and become confused (M. K. Johnson et al., 1993; Reyna & Titcomb, 1997; Titcomb & Reyna, 1995). As discussed earlier, source memory performance has been found to be independent of recognition memory, indicating the possibility that some verbatim traces remain intact, while source information does not (recognition processes, such as familiarity, would play a role here as well).

Both frameworks use the notion of source confusion errors to explain the subjective content of illusory recollections. The activation-monitoring account states that features become detached from studied items and are attached to the memory trace of a critical lure (Lampinen et al., 2005). The fuzzy-trace theory states that gist information becomes conflated with details from presented items (this is referred to as phantom recollection; Brainerd, Wright, Reyna, & Mojardin, 2001). Incorrect memories occur when attributes, including source information, become detached from the original memory, and become erroneously connected to other memories (Reyna, 2000; Reyna & Lloyd, 1997). Curiously, the fuzzy-trace theory makes few predictions regarding the possibility of source confusion errors in complex memory tasks (Lindsay & Johnson, 2000).

Another common characteristic is that both theories include an activation factor (Gallo, 2010) and a monitoring strategy (Gallo, 2004, 2010). For the activation factor, in the activation-monitoring framework it is the spreading of activation, at study and/or test, among extant conceptual representations in a mental lexicon or semantic system. Within fuzzy-trace theory activation occurs when participants mentally construct a gist representation – a summary of the common semantic characteristics or theme (Brainerd & Reyna, 1998). For the monitoring strategy, both the fuzzy-trace theory and activation-monitoring theory rely on the strategic use of recalled information to monitor the accuracy of memories (Gallo, 2004, 2010). This similarity may explain why scholars have argued that the monitoring processes defined in the two theories are currently under-developed (Gallo, 2010; Lindsay & Johnson, 2000; Reyna, 2000).

Despite the commonalities of the two theories described above, in each theory the decision process is qualitatively different. In source-monitoring various features of a memory are recalled to varying degrees, and features can be attributed to incorrect sources depending on the monitoring process being utilised (Gallo, 2006). In the fuzzy-trace theory monitoring is a recollection-rejection strategy, where verbatim traces can neutralise the ability of gist traces to create false memories (Brainerd, Reyna, Wright, & Mojardin, 2003). Using the wealth of information available about source-monitoring theories and dual-process theories (such as fuzzy-trace theory), Gallo (2004, 2006) has defined two monitoring processes believed to represent two fundamentally different underlying decision processes in which false memories can be avoided through true recollection. The two monitoring strategies are called *diagnostic* and *disqualifying monitoring*.

Diagnostic monitoring relies on expectations and refers to situations in which the absence of information reduces memory errors. That is, if remembering a questionable event failed to evoke the expected recollections for it having been previously experienced, the event would be rejected and judged as false. For instance, although an event might be plausible (e.g., "I told you I lost the house key"), the absence of recollecting expected information ("no, you did not tell me, I would have remembered if you had") can lead one to reject the event. The decision is based on the fact that expected information is missing from the memory (Gallo, 2006). Diagnostic monitoring also focuses on the quality of memory evidence for the event in question, and if the evidence passes or fails an expected criterion (Gallo, 2010). Diagnostic monitoring is an important part of source-monitoring (Gallo, 2010) because memory characteristics are believed to vary in quality and intensity (M. K. Johnson et al., 1993), and source decisions are made by comparing recollected information to the information expected to be remembered from different sources (Mitchell & Johnson, 2000).

On the other hand, Gallo (2004, 2006, 2010) describes disqualifying monitoring as a decision process that relies on collateral information. This process has been referred to as recall-to-reject (Rotello & Heit, 1999, 2000), and recollection rejection (Brainerd & Reyna, 2002b; Brainerd et al., 2003). Disqualifying monitoring takes place when the correct recollection of one event or certain information logically permits the rejection of a more questionable event as having happened. For example, "I know I did not take a taxi, because I remember taking the bus". Gallo (2010) describes the following three distinctive strategies, and notes there may be more, in which recalling information can disqualify a critical lure as having been presented during study: (a) the identify-and-reject strategy, in which one attempts to identify the critical lures and mentally tag them as such during encoding, and then at test false memories are avoided by recollecting that the item was tagged as non-presented; (b) the source-based exclusion strategy, which is when the experimental design allows a source-based exclusion rule, and (c) the exhaustive-recall-to-reject strategy which arises when a participant can recall all of the presented items, thus knowing the critical lure was not presented at study (Gallo, 2006).

Similar to difficulties with the activation-monitoring and fuzzy-trace theories, it is rarely clear whether experimental manipulations influence diagnostic monitoring, disqualifying monitoring, or both. This generality is not intrinsic to the two monitoring process models, but is because the two models are relatively new (Gallo, 2004, 2010). Gallo (2004) also believes that diagnostic monitoring could occur within disqualifying monitoring, as a successful disqualifying process may require true and false memories to be distinguished through the comparison of memorial evidence (with true memories being accompanied by more perceptual details than false memories), which is a diagnostic process. Further research into the frameworks discussed above will increase our understanding of the creation of false memories, and how the different decision processes contribute to editing false memories. One theory that further investigates age-related differences in memory performance and decision processes in the DRM paradigm is signal detection theory and its associated methodology for analysing sensitivity and response bias. The following Chapter provides an overview of signal detection theory, before discussing the two signal detection models that have been developed to explain false memories.

Chapter Three

Literature Review Part Two: Signal Detection Theory and False Memories

Although signal detection theory has existed since the early twentieth century, it has only been frequently utilised in psychological research since the publication of Green and Swets' (1966) influential classic, *Signal detection theory and psychophysics* (psychophysics is the study of the relationship between physical stimuli and their psychological experience). Now signal detection theory is employed in a number of psychological disciplines, including memory, cognition, social psychology, and non-medical diagnostics (Macmillan & Creelman, 2005). However, signal detection theory has only been utilised in DRM (Deese/Roediger-McDermott) research over the last decade or so. The purpose of early work that employed signal detection analyses to examine false recognition performance was to determine whether the patterns of effects found in the more basic accuracy analyses, were attributable to differences in sensitivity and/or response bias (e.g., Budson et al., 2000; Dodson & Schacter, 2001; Kensinger & Schacter, 1999, described below). At the same time as this early work the first signal detection models to explain false memories were developed. The introduction of these models caused controversy. The first model – the criterion-shift model theorised by Miller and Wolford (1999) – faced strong opposition from a number of scholars (e.g., Roediger & McDermott, 1999; Wickens & Hirshman, 2000; Wixted & Stretch, 2000), who proposed a second model based on the storage of items. The following sections will provide an overview of signal detection theory before discussing the criterion-shift and storage-based signal detection models for false memories, and then review DRM research that has employed signal detection theory.

Signal Detection Theory

Signal detection is a theoretical framework for describing and studying decisions when they are made under conditions of uncertainty (Macmillan & Creelman, 2005). When such decisions need to be made, in an attempt to make consistently accurate judgements individuals use rules to govern their decision process. Signal detection theory provides the means to examine individuals' responses to stimuli by calculating independent measures of response bias, the predisposition to favour one response over another, and sensitivity, the ability to detect a signal or discriminate a signal from noise (Green & Swets, 1966).

Two of the key concepts in signal detection theory are signal and noise. A signal is any stimulus or event that was presented to the observer (a research participant). Noise includes all of the stimuli in the environment which the signal needs to be detected against, and discriminated from (McNicol, 1972). For example, in the DRM paradigm, each presented list word is a signal, and lure words in the test can be considered noise. There are two forms of noise, external and internal. External noise is any type of distraction or superfluous stimulus that has the ability to disguise the strength of a signal, and thus, generate perceptual errors (i.e., causing an individual to misconstrue a stimulus) and influence an observer's sensitivity (Green & Swets, 1966; McNicol, 1972). Internal noise includes any type of brain activity (e.g., neuronal firing; Pinneo, 1966), or cognitive processes, that occurs as an individual attempts to form an accurate mental representation of the stimuli being observed (McNicol, 1972). The internal representation one forms can also be influenced by numerous observer-specific variables (McNicol, 1972). Both the mental representation one forms and decision noise can shape response bias and accuracy. This highlights the importance of measuring sensitivity and response bias (McNicol, 1972).

According to Green and Swets (1966), there are two possible states of the world, noise alone or signal plus noise. Both noise and signal plus noise states have probability distributions that provide information about the probability a perceived event or stimuli will be caused by noise or signal plus noise. In Figure 3.1 below the left distribution refers to the probability a perceptual effect, what the observer experiences on each trial, will be due to noise, and the right distribution represents the probability the perceptual effect will be caused by signal plus noise. Provided the distributions are normal with equal variances (the usual assumption), at the point where the two distributions cross it is equally probable that noise or signal plus noise caused the perceived event. On each trial an observer has to decide if a stimulus was present (signal) or not (noise). Because the distributions overlap, for some stimuli it will be unclear as to whether they reflect noise only, or signal and noise. Due to this uncertainty noise may be perceived as a signal (false alarm) or a signal may be perceived as noise (miss).

Within signal detection theory the decision observers make, and their decision rules, depend on where their criterion is placed along the continuum (Green & Swets, 1966). The criterion is a cut-off point that defines the conditions under which an event or stimulus is considered by an individual to be either a signal or noise, and tells us what an individual's response bias is (McNicol, 1972). After an individual has assumed a criterion, the rules they use to decide their response on each trial are (a) if the perceptual effect is greater than the criterion (to the right of the criterion) the item is believed to be present and the response is ‘signal’. For instance, in a DRM recognition task a participant would believe a test word was presented at study; and (b) if the perceptual effect is lower than the criterion (to the left of the criterion) the response is ‘noise’, the stimulus was not present. For example, in the DRM, a test word would be judged as new, not presented during study. As can be seen below in Figure 3.1, the ideal criterion is positioned where the two distributions intersect, a neutral criterion. A criterion to either the left or right of this neutral position can be considered bias. For example (refer to Figure 3.1), as the criterion moves to the left of the neutral position, there is a bias to respond signal (a lax or liberal criterion). Conversely, if the criterion moves to the right of the neutral position, there is a bias to respond noise (a strict or conservative criterion; Goldstein, 2010).

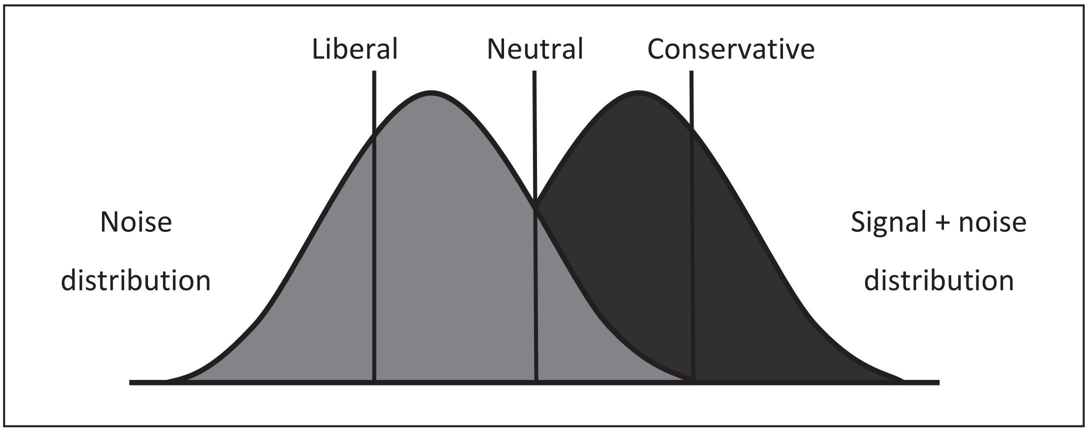


Figure 3.1. Example of noise and signal plus noise distributions and the placements of a liberal, neutral, and conservative criterion. Adapted from “Measurement of response bias in aging research,” by W. L. Danziger. In L. W. Poon (Ed.), 1980, *Aging in the 1980s: Psychological issues* (p. 553). Washington, DC: American Psychological Association. Copyright 1980 by the American Psychological Association.

The criterion observers adopt influences their hit (correctly identifying a signal as present) and false alarm (incorrectly identifying a signal as present) rates (Green & Swets, 1966; McNicol, 1972). For instance (referring to Figure 3.1 above), with the neutral criterion, only a small amount of the noise distribution falls to the right; therefore, false alarms would rarely occur. Most of the signal plus noise distribution sits to the right of the neutral criterion; therefore, hits would occur frequently. A neutral criterion is ideal, because it maximises hit rates, while minimising false alarms. Biased criteria increase either hit rates or false alarms, but at the expense of the other. For instance, if the liberal criterion was assumed, the false alarm rate and hit rate would be high, because more of the two distributions fall to the right of a liberal criterion. The use of the strict/conservative criterion would mean false alarm and hit rates would likely be low, as a smaller portion of the distributions falls to the right of the criterion.

Another reason why noise and signal plus noise distributions are important, is because the distance between the means of the distributions indicates an individual's sensitivity (Green & Swets, 1966). Sensitivity is a function of an individual's hit and false alarm rate. Someone with perfect sensitivity would have a false alarm rate of zero and a hit rate of one (Macmillan & Creelman, 2005). The most widely used sensitivity measure in signal detection theory is d-prime, denoted as *d'* (Macmillan & Creelman, 2005). It is a standardised unit of measurement similar to a *z*-score (non-parametric measures also exist). If an observer had a *d'* of zero they would not be able to differentiate signal from noise. In the DRM paradigm this would mean participants' false recognition rates would equal their correct recognition rates. Sensitivity can also be estimated by plotting the associated hit and false alarm rates as the decision criterion is moved from right (strict criterion) to left (lax criterion) through the signal plus noise and noise distributions, producing a Receiver-Operating Characteristic (ROC) curve (Green & Swets, 1966; Macmillan & Creelman, 2005). The area under the ROC curve is often used as a non-parametric sensitivity measure. Where the signal detection model assumes normal distributions with equal variances, the area under the ROC curve is equal to √2*d'* (Macmillan & Creelman, 2005).

According to Snodgrass and Corwin (1988), when applying signal detection theory to recognition memory experiments, the items on a test are often thought to lie along a strength-of-evidence, or memory strength, dimension (refer to Figure 3.2 and 3.3 below). For different item types (e.g., old or new), the probabilities of the memory strength are often assumed to be normally distributed across the memory strength dimension. Due to encoding and storing presented items, the mean level of strength is typically greater for items that were presented during study (old items) compared to items presented at test that were not presented at study (new items), as can be seen in Figure 3.2. Because the noise and signal plus noise distributions, described above, are represented by old and new items in recognition tasks, the different distributions for the different item types overlap, with the distribution for the old items placed further along the right of the dimension than that of the new items. When participants set a decision criterion along the memory strength dimension, items that fall to the left (below the criterion) are judged as new, and those that fall to the right (above the criterion) are judged as old. In recognition tasks sensitivity can be understood as the extent to which a participant can differentiate new and old items, and it is calculated by using the difference between the means of the old and new distributions. Bias is understood as the propensity for a participant to produce mainly old or new responses for 'uncertain' items, and it is calculated as the distance from where the two distributions intersect to the criterion location. Interestingly, signal detection research into age differences in recognition memory does not provide a clear cut picture of whether the differences are due to sensitivity, response bias, or both.

Age and signal detection estimates in recognition memory. Some researchers argue that age-related deficits in memory tasks reflect a bias to respond cautiously or conservatively (Botwinick, 1984; Danziger, 1980; Okun, 1976). Others believe ageing is linked to responding more liberally (Howard, Bessette-Symons, Zhang, & Hoyer, 2006; Suengas, Gallego-Largo, & Simon, 2010). The evidence is mixed. Charles, Mather, and Carstensen (2003), and Suengas et al. (2010) examined age-related recognition performance in adults for negative, positive, and neutral images. In Charles et al.'s first experiment they compared three age groups (mean ages 24.6, 46.8, and 71.0) and found that, across all the image types, as age increased the criterion moved from conservative to liberal. In their second experiment they compared younger (mean age 23.5) and older (mean age 74.1) adults and found no age difference in response bias. In both experiments the younger group(s) demonstrated greater sensitivity than the older group. On the other hand, Suengas et al. found no difference in sensitivity between younger (mean age 19.8) and older (mean age 77.3) adults, but younger adults were more conservative than older adults. Like other research, the experimental design and stimuli are probably important covariates here.

Gordon and Clark (1974b) demonstrated that when recognition for prose material was examined, older (mean age 71.2) adults had poorer sensitivity than younger (mean age 24.8) adults, but there were no age differences in estimates of response bias. Gordon and Clark (1974a) also examined recognition performance between older (mean age 71.2) and younger (mean age 24.8) adults, using word lists and nonsense syllables. They found that for recognition of the word lists, compared to younger adults, older adults had lower sensitivity and set a stricter criterion, resulting in fewer false alarms. For nonsense syllables older adults again had lower sensitivity, but adopted a more liberal criterion than younger adults. Harkins, Chapman, and Eisdorfer (1979) examined recognition performance for word lists in younger and older females (mean age 21 and 71, respectively) and found that the younger group had better sensitivity and a more conservative criterion than the older adults. Howard et al. (2006) showed that when given a picture recognition task (the pictures were nature and urban scenes of travel destinations from around the world), younger adults have greater sensitivity, and are more conservative, compared to older adults (mean age 24.4 and 71.2, respectively). By contrast, Lamont, Stewart-Williams, and Podd (2005) tested three age groups (mean age 25.9, 66.8, and 81.2) for recognition of faces and found that, the better performance (sensitivity) of the younger group, was due only to the false alarm rates of the groups. Neither hit rates nor bias were significantly different across the age groups.

Poon and Fozard (1980), and Breck and Baron (1987) examined age-related differences in continuous recognition tasks, and found different explanations for older adults' poorer performance, compared to younger adults, using signal detection analyses. Poon and Fozard presented male participants, from three different age groups (median ages were 20, 52, and 63), with a list of words, where some words were repeated during the presentation and others were not. After being presented with each word the men had to decide if it was old (previously presented in the list) or new. Their results illustrated that across the three age groups recognition performance was similar, but the oldest group had lower sensitivity and a more liberal response bias than those in the younger age groups. Breck and Baron presented younger (18-26 year olds) and older (62-75 year olds) women with continuous lists of nonsense letter-number combinations (e.g., B39H), with repeated items being presented at intervals ranging from 0-32 items. Contrary to their expectation, there was no difference between the younger and older groups in their response bias estimates, but there was a significance difference between the groups in their estimates of sensitivity. The younger group had greater sensitivity than the older group.

Signal Detection Models of False Recognition

The previous section has provided an overview of signal detection theory and its application to recognition memory, including a review of research investigating ageing and signal detection estimates of recognition memory. Even from the short review above, it is clear that age-related differences in recognition tasks do not conform to a single pattern of signal detection results. The same can be said for signal detection analyses in false memory tasks, discussed below. The current section focuses on the two false memory models: the criterion-shift model and the storage-based model.

The criterion-shift model. Miller and Wolford (1999) proposed that item-to-item criterion shifts (refer to Figure 3.2 below) explain data patterns in false memory experiments. To come to this conclusion Miller and Wolford conducted DRM research in which the recognition test contained (a) critical lures that were not presented at study, (b) critical lures that were presented within their DRM list at study, (c) DRM list items that were presented at study, (d) DRM list items that were not presented at study, (e) unrelated items that were presented at study, and (f) unrelated items that were not presented at study. As a result, independent hit and false alarm rates for the three item types (critical lures, DRM list items, and unrelated items) could be calculated. In two experiments they found that for the different item types estimates of sensitivity were similar, whereas there were large differences between estimates of bias. Miller and Wolford argued that the similar sensitivity estimates meant each item type profited equally from being presented during study. The research also found false alarm rates for critical lures were the highest, followed by related items, then unrelated items. Miller and Wolford argued these results were due to shifts towards a more lax criterion across the item types, with critical lures having the most lax criterion, followed by related then unrelated items.

In Miller and Wolford's (1999) criterion-shift model, false recognition is due to holding a different criterion for each item type (as can be seen in Figure 3.2). False alarms do not represent 'real false memories' based on either recollection or familiarity. Rather, participants strategically infer, using meta-knowledge from experience with the task that related words they cannot remember were, probably, presented (Gallo, 2006). The meta-knowledge participants develop as the task proceeds is thought to concern the structure of the lists (Miller & Wolford, 1999); that is, participants adopt a criterion that considers the apparent likelihood the item is a member of a list that was studied (Wickens & Hirshman, 2000). When a participant recognises a lure is related to the list it leads them to employ a more liberal criterion, and respond 'old' more often. Because of the way the list is constructed, critical lures have the highest probability of being recognised as related to the list, followed by related items and then unrelated items (Miller & Wolford, 1999). The storage portion of the model assumes that for all unstudied items the strength of evidence that items are old is taken from the same distribution. Presenting items adds to its strength, thus shifting the distribution mean, and possibly altering the variance (Wickens & Hirshman, 2000). Activation is assumed to influence the placement of the decision criterion, but not the extent to which an individual remembers previously encountering an item (i.e., strength of evidence). Therefore, strong associates of list items will have a liberal criterion (Wixted & Stretch, 2000).

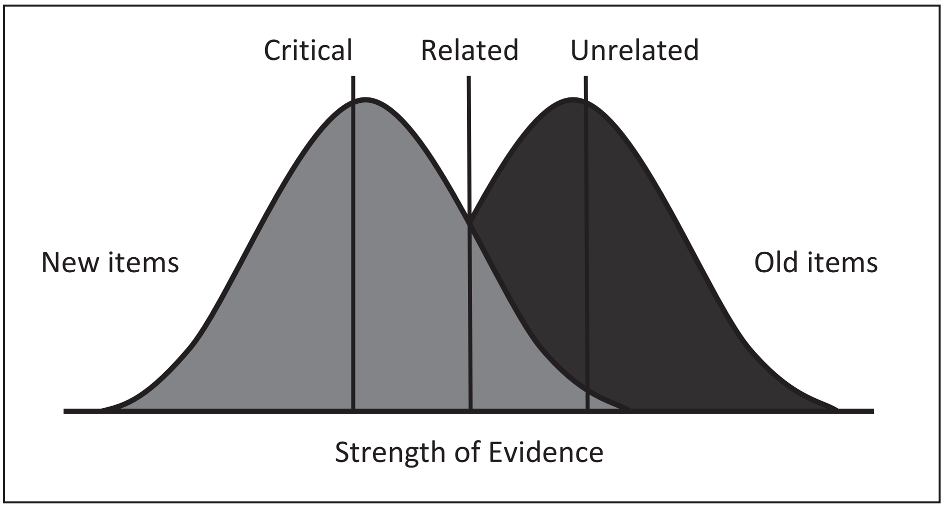


Figure 3.2. Hypothetical distributions (old and new items) and criteria (critical lures, related items, and unrelated items) according to Miller and Wolford's (1999) criterion-shift model. Adapted from "The Case Against a Criterion-Shift Account of False Memory," by J. T Wixted and V. Stretch, 2000, *Psychological Review, 107,* pg. 372. Copyright 2000 by the American Psychological Association.

Miller, Guerin, and Wolford (2011) have published research to clarify the original criterion-shift model, stating that it assumes participants adopt two underlying criteria for recognition judgements (not three). A liberal one for any item presented at test that is perceived to be thematically related to studied lists, and a conservative criterion for the items that do not appear to be thematically related to studied lists. However, using their model analyses will result in three criteria because the criterion for each of the item types is a product of different probability mixtures from the two underlying criteria.

The storage-based model. The introduction of the criterion-shift model triggered strong debate, and produced in-depth theoretical discussion concerning the nature of false memory phenomena. Roediger and McDermott (1999) argued against the criterion-shift model, stating that it was implausible for a number of reasons. To contend with the criterion-shift model Wixted and Stretch (2000) developed the storage-based signal detection model, which was derived from Roediger and McDermott's false-memory model. One of the model's assumptions is that participants set a single criterion and examine each test item against this set criterion (refer to Figure 3.3 below). Another fundamental assumption is that recognition decisions are based on a one-dimensional strength of evidence continuum. At test, the strength (S_i) of an item is thought to be a function of the direct effects from the item being presented (P_i), and the additional indirect effects owing to the associative activation (A_i) from other presented list items. This simple model ($S_i = P_i + A_i$) fits with a number of more specific theoretical assumptions regarding how memory strength for items increases. For example, strength from P_i could be from conscious rehearsal of presented items or enhanced perceptual fluency, but it is a quantity of strength that only exists if the item is presented. For lures, P_i would equal zero. Strength from A_i may be due to rehearsal of an activated item, or simply due to the unconscious spread of activation, but it only exists because an item was associatively activated (Wixted & Stretch, 2000).

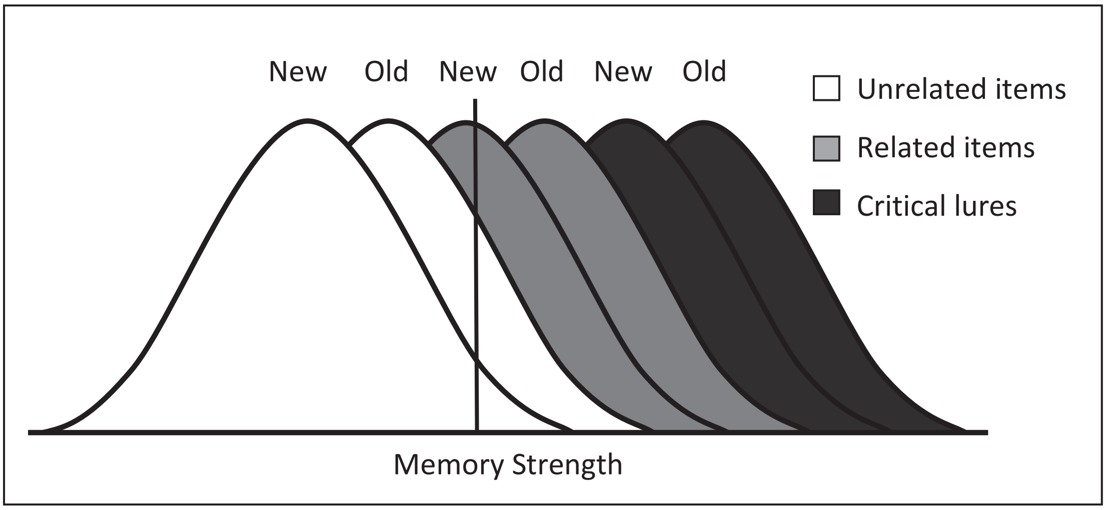


Figure 3.3. Hypothetical distributions (new unrelated items, old unrelated items, new related items (e.g., DRM list items), old related items, new critical lures, and old critical lures, respectively) and a single criteria according to the storage-based model (e.g., Roediger & McDermott, 1999; Wickens & Hirshman, 2000; Wixted & Stretch, 2000). Adapted from "Sensitivity Reductions in False Recognition: A Measure of False Memories with Stronger Theoretical Implications," by C. E. Westerberg and C. J. Marsolek, 2003, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* pg. 749. Copyright 2003 by the American Psychological Association.

In the storage-based model the initial strength associated with an item is based on a common distribution, and presenting an item enhances the memorial evidence that the item is old; thus, studied items have a different distribution to unstudied items. Presented items can impact the strength of categorically- or semantically-related items. Consequently, for category and semantic members of a list the evidence distribution shifts farther to the right, whether they are presented or not (Wickens & Hirshman, 2000). One mechanism thought to produce this distribution shift, in line with the spreading activation model, is that the presentation of items causes information about those items and *related items* (presented *and* non-presented) to be stored. An alternative possibility is that for unstudied items (related and unrelated) no memory of them is directly stored, but during recognition when memory for the item is probed, any related item that has been stored in memory may generate a signal for the unstudied items due to their shared attributes, resulting in an 'old' response for an unstudied item (Wickens & Hirshman, 2000). In the storage-based model, when calculating the signal detection parameters, differences in bias will result due to the differences in memory strength of the three types of items. Items with stronger memory evidence have their distributions farther to the right; therefore, the distance between the criterion and the associated distribution differs (Westerberg & Marsolek, 2003; Wixted & Stretch, 2000). Because of the nature of the list, the critical lure will be the most strongly related and activated word associated to the presented items, moving the critical lure distribution further to the right, and thereby generating an apparently liberal response bias (Westerberg & Marsolek, 2003).

Research into the two models. With the inception of the two signal detection models researchers have attempted to establish which model best explains that data. This is no easy feat, because even though the two models generate very different explanations for false memories, their signal detection parameters are the same, and calculating the parameters will not provide the information necessary to differentiate the models (Miller et al., 2011; Westerberg & Marsolek, 2003; Wickens & Hirshman, 2000; Wixted & Stretch, 2000). This is the case because taking into account signal detection theory assumptions it is correct that a change in decision strategy necessitates a change in the measured bias. It does not logically follow that only a change in response bias (as found by Miller & Wolford, 1999) causes a change in the decision strategy; changes in memory distributions could lead to changes in bias. When items differ in their familiarity, their distributions also differ. When the criterion is fixed across the different DRM test items, as in the storage-based model described above, the distance between the criterion and each type of items distribution differs, and produces different bias estimates (Wixted & Stretch, 2000). To examine the models, researchers have had to manipulate the false memory effect by using different variables, such as presentation duration, and then inspect how each models' parameters alter with the manipulations.

Stretch and Wixted (1998) conducted research attempting to cause participants to use item-by-item criterion shifts based on metaknowledge. To do this, in two experiments[6], they continuously manipulated the strength of items within lists. In a list of studied words half of the words were green and presented once (weak), the other half were red and were presented five times (strong). The recognition test contained studied words in the colour they were presented in and lures that were either green or red. Stretch and Wixted believed that differences between false alarm rates for red and green lures would provide evidence of a criterion shift, because a stringent criterion could be employed to avoid red lures. By contrast, participants would lower their criterion for green items to avoid missing targets, based on their knowledge that a green word might not feel as familiar as the red words, because green words were not repeated during study. If the participants used an item-by-item criterion shift the rate of green false alarms would be higher than those for red. Their results showed no evidence for an item-to-item criterion shift; false alarms for the red and green words were almost identical in both experiments.

Wickens and Hirshman (2000) used data from an experiment by Arndt and Hirshman (1998) in which DRM lists were presented using durations of 300, 800, or 3,000 ms. Participants were presented with a 16-item DRM list using one of the presentation durations, and then completed a brief distracter task before a recognition test. The recognition test contained four different types of items: presented words, critical lures, unrelated words, and new critical lures (critical lures from DRM lists that were not presented). It was found that the longer the presentation duration the more presented items and critical lures were recognised, while fewer unrelated words and new critical lures

[6] The two experiments were exactly the same, except that in the second experiment (which in the article is Experiment 5) participants were explicitly told that the red items would be presented five times. In the prior experiment (Experiment 4) participants were told that some items would be strengthened using repeated presentation, but participants were not told strong and weak items would be presented in different colours.

were recognised (Arndt & Hirshman, 1998). Wickens and Hirshman fitted the two signal detection models to the data and examined the parameter values. The criterion-shift model included three very similar distributions (unrelated words, critical lures, and presented words), and the use of two decision criteria (one for presented items, and one for non-studied items). For the storage-based model each of the four item types had its own distribution – with each placed in a different position relative to memory strength – and there was a single criterion for all decisions. The obtained parameter estimates showed that for both models the distribution mean and criterion for presented items were equal in each of the three presentation durations. For the longest presentation time the related item distribution mean and criterion were greater than for the shorter presentation times, which were the same. Wickens and Hirshman believe that the wider distribution spread, which is due to stronger memories gained from longer presentations, caused a shift in criterion so it was more optimally placed. For the critical lures the parameters associated with the two models differed. Under the storage-based model the distribution mean shifted due to the presentation of related items, and this shift was greater in the longest presentation duration compared to the shorter times. This is not surprising, as increasing the strength by presenting items longer also increases the associated strength of the critical lures. The increase in strength for the critical lures was less than that for the studied items from the shortest to longest duration. For the criterion-shift model the criterion for critical lures shifts to be more lax as presentation duration increases. Wickens and Hirshman stated that this shift was substantial, representing a tremendous degree of change in bias. In fact, if the difference was due to a criterion shift and nothing else, then participants must have believed that the critical lures in the shortest presentation duration were more likely presented than the presented DRM list items. The criterion-shift model would, somehow, have to adequately argue that with the longer presentation, participants became aware of the special nature of the lure words, and the increased presentation duration caused participants to respond to the critical lures quite differently from the related items (Wickens & Hirshman, 2000). Overall, Wickens and Hirshman believed their results were somewhat better explained by the storage-based model over the criterion-shift model.

Westerberg and Marsolek (2003) investigated the two signal detection models by performing three experiments, which were specifically designed to examine the possibility that there are differences in sensitivity for DRM critical lures, presented words, and unrelated words. Thus far, research had not focused on differences in sensitivity estimates. In Experiment 1 each participant heard a number of DRM lists and two lists of unrelated words. The recognition test contained old and new critical words, old and new related words, and old and new unrelated words. Hence, independent measures of bias and sensitivity for each word type could be calculated. Westerberg and Marsolek's results demonstrated that sensitivity for related words was significantly greater than sensitivity for critical lures and unrelated words. Unexpectedly, the sensitivity of critical lures and unrelated words was comparable. For the bias estimates there was a significantly greater bias to respond old to critical lures compared to related words, and a significantly greater bias to respond old for related words compared to unrelated words.

Due to the significant bias differences found in Experiment 1, Westerberg and Marsolek (2003) conducted a second experiment that employed a procedure to prevent these differences. The procedure was a two-alternative forced-choice task and each pair (one old item and one new item) presented at test was two critical lures, two related words, or two unrelated words. That is, participants were presented with multiple pairs of items, both of which were the same type of item, and were forced to recognise one as new and one as old. Westerberg and Marsolek believed this procedure would attenuate or possibly eliminate the large bias found in Experiment 1, but still allow sensitivity differences to be observed, potentially demonstrating that bias effects are not linked to differences in sensitivity. As expected, the results showed the bias measure was near zero for all of the items, with no significant differences. Sensitivity of critical lures was significantly less than that for related and unrelated words. Sensitivity did not differ for related and unrelated items. These results indicate that bias differences are not required for changes in sensitivity for critical lures.

Because Experiment 2 showed no sensitivity difference between unrelated and related words, and both signal detection models predict sensitivity will be lowest for critical lures, followed by related words, and then unrelated words, a third experiment was performed. Westerberg and Marsolek (2003) equated the related and unrelated test words on their frequency of occurrence in the English language – in Experiments 1 and 2 frequencies were equated for unrelated and critical test words only – because they believed that this factor may have caused the unrelated and related items' sensitivity to be equal. The procedure was the same as in their first experiment. The results showed that bias estimates were significantly different between each item type, with the critical lures having the most liberal bias, followed by related words, and then unrelated words. Critical lure sensitivity was greater than both related and unrelated word sensitivity, but sensitivity did not differ between related and unrelated words. The patterns of results found in these three experiments are not predicted by either the storage-based model or the criterion-shift model, indicating that both models need review and elaboration (Westerberg & Marsolek, 2003).

DRM Research with Signal Detection Analyses

With the controversial introduction of the signal detection models it is unsurprising that at the present time few DRM research articles employ signal detection analyses (relative to what does exist for the DRM paradigm), and those that do, place the signal detection results secondary to the more basic analyses of recognition accuracy (e.g., Benjamin, 2001; Hicks & Marsh, 2001; Waldie & See, 2003). However, before the two models were introduced, researchers (Koutstaal & Schacter, 1997) had developed a method to use signal detection analyses to examine item-specific memory and gist memory in false recognition paradigms. In this research 'gist' is based on Reyna and Brainerd's (1995a) gist memory trace in the fuzzy-trace theory, and the idea of the general similarity of information between items (Curran, Schacter, Norman, & Galluccio, 1997; Hintzman & Curran, 1994). Research employing the gist and item-specific signal detection analyses are discussed next, followed by research that has employed more classical signal detection analyses to supplement basic accuracy analyses.

Signal detection analyses of item-specific and gist memory. Koutstaal and Schacter (1997) developed signal detection analyses that examine item-specific and gist memory in false recognition paradigms. Although their research used pictures and not a DRM task, it will be discussed here because it was the first to use the type of signal detection analysis discussed here, and since its publication the analysis has been utilised by others to examine DRM performance. In Koutstaal and Schacter's signal detection analyses item-specific memory is examined by comparing hits to false alarms of unrelated pictures and hits to false alarms of related pictures. In the signal detection analyses for gist memory, because false alarms to related lures are thought to be due to gist memories, they are treated as hits (and are referred to as 'gist hits'). Therefore, related false alarms (gist hits) are compared to unrelated false alarms.

Using the previously described analyses, Koutstaal and Schacter (1997) investigated older (mean age 68.7) and younger (mean age 18.8) adults' recognition of pictures. Participants were shown numerous pictures that fit into one category (e.g., different boat pictures), and then at test were asked to distinguish pictures that were presented from pictures that were not. Non-presented pictures were either from the same category as the presented pictures, or an unrelated category. They used pictures because they believed recognition errors would be based on the similarity (gist) of items, not source errors, as the task would not encourage the creation of the specific false test items. Across three similar experiments younger adults consistently showed greater correct recognition and lower false recognition than older adults.

In the signal detection analyses younger adults consistently demonstrated greater sensitivity for hits compared to false alarms, which were believed to measure item-specific memory, and used a more conservative criterion, than older adults. For related item false alarms (gist hits) compared to unrelated false alarms, which was believed to measure gist memory, older adults always had greater sensitivity, and a more liberal bias than younger adults. Koutstaal and Schacter believed this result indicated older adults employed gist representations whereas younger adults did not. To examine if sensitivity, bias, or both were responsible for age-related differences in false recognition, an analysis was conducted in which age groups false alarm rates for unrelated items were matched. The analysis indicated older adults' willingness to use gist representations and that differences in response bias were responsible for their higher levels of false recognition compared to younger adults. According to Koutstaal and Schacter, these results fit the fuzzy-trace theory, because younger adults' hits relied on item-specific memories. On the other hand, older adults' hits and false alarms depended on the use of gist memories.

Kensinger and Schacter (1999; described earlier) utilised the signal detection analyses developed by Koutstaal and Schacter (1997) examining age-related differences in a DRM task with five study-test trials. To recap, false recognition rates decreased across the trials for younger adults, but older adults' (mean age 19.3 and 68.2, respectively) false recognition rates remained the same. Both younger and older adults' true recognition rates increased across the trials. The signal detection analyses for Kensinger and Schacter's research determined that for item-specific memory (hits compared to false alarms for unrelated items) older and younger adults had equal sensitivity on the first trial and both age groups increased their sensitivity across trials. There were no age differences in bias across all five trials, though both age groups adopted a slightly more liberal criterion across the trials.

The second set of analyses for item-specific memory (hits compared to related false alarms) indicated that on the first trial participants could not distinguish correct from false items, as sensitivity was almost at chance level for both age groups, .53 for younger adults and .57 for older adults. By the final trial both age groups had increased their sensitivity, the younger adults more than the older adults, and become more conservative, with older adults responding more liberally than the younger adults. The analyses for gist memory (related item false alarms, i.e., ‘gist hits’, compared to unrelated false alarms) indicated that on the first trial younger adults were more sensitive than older adults, but over trials older adults’ sensitivity increased whereas younger adults’ sensitivity decreased. Also, both age groups became increasingly conservative across trials, and the trend was more prominent for the younger adults. Similar results, using five trials, have been found by Budson et al., 2000. These results suggest that older adults’ increase in false memories stems from both an increase in their willingness to rely on gist traces *and* a more liberal response criterion.

More recently Dodson and Schacter (2001) utilised Koutstaal and Schacter’s (1997) signal detection analyses to examine the contribution of decision processes when rejecting false memories. In their first experiment participants studied 16 15-item DRM lists which were either presented visually and aurally (seen and heard), or visually with the participants also speaking the word aloud (seen and spoken). Both of these conditions resulted in equal recognition rates for studied items and unrelated words. Conversely, those in the *seen and spoken* condition had lower false recognition scores than those in the *seen and heard* condition. In the signal detection analyses for hits compared to unrelated false alarms (item-specific memory) sensitivity and bias were equal for the two conditions. In the analysis for hits compared to related false alarms, the *seen and spoken* condition resulted in higher sensitivity (.69) than the *seen and heard* condition (.50, chance level). When examining gist memories (related item false alarms compared to unrelated false alarms) the *seen and heard* condition had a greater sensitivity score, along with a more liberal response bias, than the *seen and spoken* condition. According to Dodson and Schacter their results indicated that suppression of false memories in the *seen and spoken* condition was due to more distinctive information being encoded during study. At test the absence of memory information for speaking an item signified that it was not presented during study.

Signal detection analyses as supplementary analyses. As described earlier, Hicks and Marsh (2001) performed three experiments. In each experiment DRM lists were presented by two different sources: heard or generated via anagrams, heard or seen, and female or male voice. At test participants completed a standard old/new recognition test only, or the standard recognition test with an added source judgement task. In all three experiments false recognition was higher when participants were asked to judge the source of items compared to when they were not. To gain a better perspective on the recognition results, Hicks and Marsh calculated signal detection measures for the discriminability (sensitivity) of old and new items, and bias (the measures used false alarm rates for unrelated items, not critical lures). In Experiment 1 (heard or anagrams), overall there was no difference in discriminability between the standard old/new test and the test with the added source judgement. However, in both conditions the words presented through anagrams were more memorable (had a higher sensitivity estimate) than items heard. In the source judgement condition participants had a liberal bias. In the recognition-only condition bias was conservative. For Experiment 2A (heard or seen items), no significant differences in discriminability were found in the source condition. In the recognition-only condition, items that had been seen by participants were better discriminated than items that had been heard. In both test conditions of Experiment 2B (female versus male voice), discriminability was equal. In Experiment 2A and 2B bias was significantly more liberal in the source condition compared to the recognition only condition. The signal detection results found in Hicks and Marsh's (2001) research are contradictory to the expectation that source memory instructions should lead participants to more carefully examine their memories. Hicks and Marsh believed that in the source test the weight placed on retrieving information to specify the source of items may have made participants assess the occurrence of items less meticulously, causing the use of a more liberal criterion.

Similar to the research above Benjamin (2001) employed signal detection analyses to supplement basic accuracy analyses in a DRM task. In this research younger (mean age 22.4) and older (mean age 74.3) adults were presented with DRM lists either once or three times. When lists were presented three times older adults had higher false recognition whereas younger adults was lower compared to when DRM lists were presented once. Both age groups increased true recognition with three compared to one presentation. The signal detection estimates showed that with repetition, the younger adults' criteria and sensitivity increased. By contrast, older adults had no change in their sensitivity, but moved towards a more liberal criterion. Older adults' increased true recognition with repetition was due to the adoption of a more liberal criterion. These results suggest that older adults maintain the automatic influence of familiarity or associative activation, but lack the memorial ability to improve their criterion; that is, adopt a criterion that maximises hits and minimises false alarms, with repetition (research by Dodson and Schacter, 2002, has shown that by using a distinctiveness heuristic older adults are able to decrease false recognition of pictures and words, compared to words only, by improving their criterion. However, the research did not use the DRM paradigm).

Waldie and See (2003) presented younger and older participants with six 12-item DRM lists, each followed by immediate recall, and then a final recognition task. False alarm scores were calculated by adding the false recognition scores of critical lures, weak lure words, and unrelated words. Their results showed that older adults recalled and recognised fewer presented items and more critical and weak lures. Their signal detection analyses showed that younger adults were more sensitive than older adults, but response bias was equal across age groups. Thus, unlike Benjamin's (2001) findings, older adults' increased false recognition and decreased correct recognition was related to their inability to discriminate old and new items, not because of a lower criterion.

Lastly, a recent study (Jou, 2011) was conducted to examine if participant confidence ratings in a DRM recognition task reflect only a criterion shift, or both a criterion shift and a change in sensitivity. Jou (2011) had participants study 12 15-item DRM lists each followed by a short distraction task before completion of a recognition test in which participants responded old/new and rated their confidence. The results revealed a false alarm rate of .71 and liberal response bias, with negative bias estimates across all confidence levels. Also, bias and sensitivity estimates both increased as participants' confidence ratings increased. These results indicated that higher confidence ratings were associated with both a shift in the criterion participants employ, and increases in sensitivity estimates. Jou believed his results indicated that participants have a conscious sense of the difference in qualities between true and false memories, and can increase or decrease their decision criterion to alter recognition performance.

Overall, signal detection theory has been utilised relatively seldom in DRM research. It seems that this limited application has produced mixed results. This inconsistency has likely been exacerbated by signal detection analyses being used in different ways across studies, and that the individual DRM experiments have differed substantially. That is, the lack of more accurate replications across experiments means results cannot be easily compared. Nonetheless, signal detection theory may yet provide a way to gain further understanding of false memory performance, particularly when comparing age groups.

Summary

This literature review has demonstrated that false memories are a robust and well researched phenomenon. Research examining performance in false memory tasks across age groups has also been thoroughly examined. It appears that older adults typically perform more poorly than younger adults, but not all research finds age-related differences. Additionally, theories describing the underlying mechanism of false memories have been rigorously examined (the most accepted being the activation-monitoring theory and the fuzzy-trace theory). These theories are both reasonably well established and can explain age-related performance differences in a variety of false memory tasks. Unfortunately, these two theories have a number of common qualities that make it difficult to design research, or find results, that support one of these theories but not the other. The third theoretical approach that was discussed was signal detection. Compared to the activation-monitoring theory and the fuzzy-trace theory, explanations of false memories and aged-related differences based on signal detection theory are relatively new to DRM research. Employing signal detection analyses may provide valuable information about younger and older adults' ability to discriminate true from false items and how they make old/new judgements.

The current study contributes to the existing research literature by investigating the creation of false memories between younger (16-30 years) and older (75-80 years) adults. The aim is to examine performance differences between the two age groups, as well as attempting to establish if age-related differences in false recognition are best explained by one of the two most accepted, but opposing, theories (the fuzzy-trace theory or the activation-monitoring theory), or if the differences can be explained better as a product of sensitivity and/or criterion differences in signal detection ability. The false memory task is based on the DRM paradigm, using the presentation of multiple DRM word lists, and a recognition task that includes weak lures, believed to provide a more sensitive measure of false recognition than if only critical lures and unrelated items are used. A source-monitoring manipulation is also employed. Participants assigned to the source-monitoring test condition will be instructed to monitor and judge the source of each test item, while participants assigned to the standard test condition will only make an old/new recognition judgement. This research is important because it will employ a novel combination of procedures (source-monitoring instructions, weak lure measures) and analyses (signal detection theory) in an attempt to deepen our current understanding of the creation of false memories and age-related effects. Noteworthy, due to the exploratory nature of this project and the theories being under-developed (with common characteristics) it is difficult to make detailed predictions for the present research.

Chapter Four
Method

The aim of the present research was to investigate age-related differences in false recognition and the theoretical approaches used to explain false memory creation. The present study was a mixed between, within subjects (2 x 2) experimental design. One of the independent variables was Age (comparing performance of 16-30 year olds with 75-80 year olds) and the other independent variable was the pen-and-paper Deese (1959) Roediger and McDermott (1995; DRM) recognition task participants were assigned to. The two recognition tasks had the same test-items, but one version was created as a standard recognition task and the other version was developed to encourage participants to use a source-monitoring strategy. The four dependent variables measured were recognition of the different types of items in the recognition test - critical lures, weak lures, unrelated words, and presented items – each of these items and the two recognition tests are described below.

Participants

Overall there were 88 participants, 41 younger and 47 older adults. However, data from 8 individuals were excluded. One younger adult scored below 26 on the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005), one older adult did not complete the DRM task, and six older adults scored below 24 on the MoCA (MoCA cut-off scores will be discussed further below). These participants are excluded from any further discussion or analyses. After the exclusions the sample consisted of 80 participants, 40 younger and 40 older adults. The younger group consisted of 25 females and 15 males, and their mean age was 22.83 years (*SD* = 4.07, range: 16-30 years). The older group consisted of 23 females and 17 males; their mean age was 77.25 (*SD* = 1.61, range: 75-80 years). As can be seen below in Figure 4.1, the levels of education differed between the age groups. Many (25%) of the older adults had fewer than 11 years of formal education, and thus any formal qualification (e.g., NCEA Level 1 or 5th Form School Certificate). By contrast many (35%) of the younger adults completed secondary school (13 years of formal education). The number of younger and older adults who completed formal training after school was 7.5% and 15%, respectively. The number of younger and older adults who held undergraduate degrees was similar, 20% and 22.5%, respectively. None of the older adults held post-graduate degrees, whereas 20%

of the younger adults did. Differences in levels of education were expected due to comparing today's education – availability and societal expectations – to that of the middle of the twentieth century. Also, some differences were likely a function of the younger adults being mainly university students.

All participants were asked to use their hearing and reading aids if they required them. Participants were volunteers, recruited through community groups, advertisements, and by word of mouth, from the Palmerston North and Marlborough wider regions.

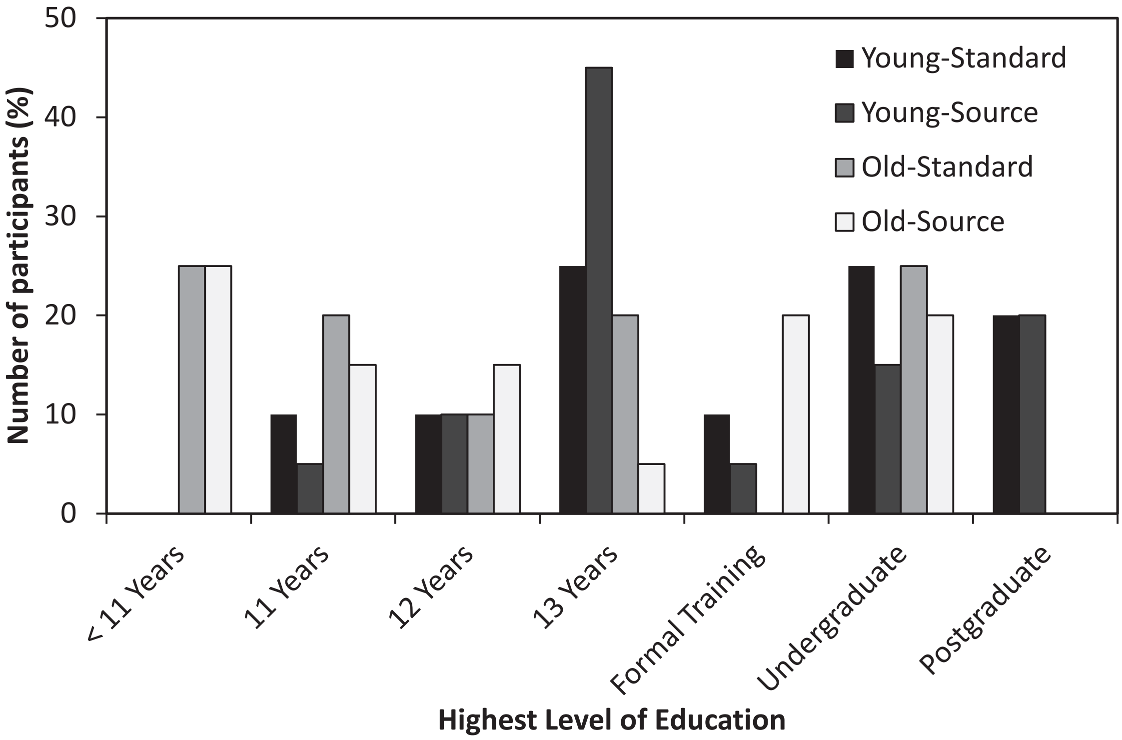


Figure 4.1. Percent of participants (per age and recognition test group) who held each level of education (identified as their highest). Eleven years of education includes NCEA Level One and 5th Form Certificate. Twelve years of education includes NCEA Level two and 6th Form Certificate. Thirteen years of education includes NCEA Level three, 7th Form (higher school) Certificate, and Bursary. Formal training is any type of formal studying and training, including, at a technical institute, as an apprentice, and registered professions that required studying and training (e.g., nurses, mid-wives).

Measures

The false recognition task. This measure utilised the DRM paradigm. It consisted of a single study phase and a single recognition test that had two versions. Each version consisted of the same items, but the instructions and response options differed. To create the task, eight DRM lists, each consisting of 15 words, were employed. Table 4.1, below, displays the DRM list items and their lures. The eight DRM lists chosen corresponded to the following critical lures: *sleep, rough, slow, mountain, city, black, fruit,* and *lion.* Six lists were taken from Roediger and McDermott's (1995) original 24 lists, and two (*city* and *lion*) originated from McDermott's (1995) unpublished materials and were printed in Stadler et al.'s (1999) article. The eight DRM lists were chosen based on previous research (Sim, 2010b) and DRM normative data (Stadler et al., 1999). The DRM lists for the lures *city*, *mountain*, *rough*, *sleep*, and *slow* were employed in previous false recognition research that examined age-related differences and was conducted with a New Zealand population (Sim, 2010b). Of these five lists, *city*, *rough*, *sleep*, and *slow* showed the greatest difference in false recognition (.17 to .12) of the critical lure between the younger and older adults, and had overall rates of false recognition that were medium-high (.69, .83, .89, and .66, respectively; Sim, 2010a). The *mountain* list was chosen because in Sim's (2010a) research it had a medium false recognition rate (.55), and was the list with the lowest false recognition rate in the study. The final three lists (*black*, *fruit*, and *lion*) were chosen as they were amongst those that had the lowest levels of false recognition (.49, .45, and .33, respectively) in the norms created by Stadler et al. (1999). They were also preferred over other lists with low levels of false recognition because the list items seemed less culturally specific. For example, the list *king* had the lowest false recognition rate (.27), but was excluded due to the norms being American-based and America having a weaker relationship with the British Monarch than New Zealand. Therefore, with a New Zealand sample, it is possible that false recognition for the *king* DRM list could be much higher.

In an attempt to control for order effects a balanced Latin square was employed to order the presentation of DRM lists to participants. This ordering strategy ensured that each list appeared before and after each other DRM list an equal number of times, and that in the study phase each list had the opportunity to be presented at the start, middle, and end position. Table 4.2 below demonstrates the eight list presentation conditions created with the balanced Latin square algorithm. Participants were randomly assigned to one of the order conditions, with all conditions being used an approximately equal number of times in each of the four age and condition groups.

As stated above, each version of the recognition test had the same items. Each test consisted of 48 presented words, eight critical lures, 24 weak lures, and 16 unrelated words. The presented words were six words from each of the eight presented DRM lists, taken from positions 2, 4, 7, 9, 12, and 14 (two from the beginning, middle, and end of each list). The critical lures were those that corresponded to the presented DRM lists. The unrelated words were chosen from DRM lists not used in the current research, with only one word being used from each DRM list, to avoid false recognition of unrelated words purely due to an associated word also being present. To choose the unrelated words, 16 DRM lists were drawn out of a bag that contained 28 lists. Then one word from each list was, again, chosen from a bag. To ensure these unrelated words were indeed unrelated to the presented DRM lists, the word association norms were consulted (Jenkins, 1970). None of the unrelated words were in the norm lists that the presented DRM lists were created from.

Weak lures were included in the test to provide a more sensitive false recognition measure. To create the weak lures for the DRM tasks, the current research used a method similar to Roediger and Gallo (1995), who found 21% of weak lures were classified as old by participants. Weak lures (three words for each DRM list) were chosen from their corresponding word association norms (Jenkins, 1970). The weak lures chosen were the closest associates to the critical lure that were not already part of the corresponding DRM list, and were not strongly semantically related to a DRM list other than the intended list (again, examined by using Jenkins, 1970, word association norms). The final weak lures (Table 4.1) came from positions 2-14 (the lower the number the stronger the word is semantically associated to the critical lure) in their corresponding norms. Finally, to decide the order of the words in the test, all of the words were placed in a bag and picked out at random.

There were two versions of the recognition test (participants were randomly assigned to one or the other). One was a standard old/new recognition test (Appendix B) and the other was created to encourage participants to use a source-monitoring strategy (Appendix C). In the old/new (standard) recognition test, the response options were: *I heard the word* or *I did not hear the word* (heard/not heard). In the source-monitoring test the response options were: *I heard this word*, *I heard this word and thought of it on my own*, *I did not hear this word but thought of it on my own*, or *I did not hear or think of this word*. These source-monitoring options were based on research conducted by Multhaup and Conner (2002). Both versions of the recognition test contained the same 6-point rating scale for how certain the participant was that the word was heard or not. The six ratings were:

very certain heard, *mostly certain heard*, *fairly certain heard*, *fairly certain not heard*, *mostly certain not heard*, and *very certain not heard*.

Table 4.1

DRM Critical Lures, Weak Lures, and List Items

Critical Lure	Weak Lures	List Items
Sleep	Comfort, sound, pillow	Bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, nap, peace, yawn, drowsy
Rough	Soft, hands, wood	Smooth, bumpy, road, tough, sandpaper, jagged, ready, coarse, uneven, riders, rugged, sand, boards, ground, gravel
Slow	Sign, go, lazy	Fast, lethargic, stop, listless, snail, cautious, delay, traffic, turtle, hesitant, speed, quick, sluggish, wait, molasses
Mountain	High, snow, stream	Hill, valley, climb, summit, top, molehill, peak, plain, glacier, goat, bike, climber, range, steep, ski
City	Square, people, building	Town, crowded, state, capital, streets, subway, country, New York, village, metropolis, big, Chicago, suburb, county, urban
Black	Sheep, red, dog	White, dark, cat, charred, night, funeral, colour, grief, blue, death, ink, bottom, coal, brown, gray
Fruit	Fly, cake, food	Apple, vegetable, orange, kiwi, citrus, ripe, pear, banana, berry, cherry, basket, juice, salad, bowl, cocktail
Lion	Lamb, zoo, mouse	Tiger, circus, jungle, tamer, den, cub, Africa, mane, cage, feline, roar, fierce, bear, hunt, pride

Table 4.2

The Eight DRM List Presentation Conditions, Ordered Using the Balanced Latin Square Algorithm

Condition	DRM list presentation order
1	lion, mountain, rough, sleep, fruit, city, black, slow
2	mountain, rough, sleep, fruit, city, black, slow, lion
3	slow, lion, mountain, rough, sleep, fruit, city, black
4	rough, sleep, fruit, city, black, slow, lion, mountain
5	black, slow, lion, mountain, rough, sleep, fruit, city
6	sleep, fruit, city, black, slow, lion, mountain, rough
7	city, black, slow, lion, mountain, rough, sleep, fruit
8	fruit, city, black, slow, lion, mountain, rough, sleep

The Montreal Cognitive Assessment. The Montreal Cognitive Assessment (MoCA) is a 12-item screening tool for mild cognitive impairment. It is a standardised test, completed individually, and takes approximately 10 minutes to administer (Nasreddine et al., 2005). The original English version of the MoCA (see Appendix D for a copy of the test and administration instructions) was used in the present research to screen all participants for possible cognitive problems. A cognitive screening measure was employed because prior research indicates that older individuals with cognitive problems exhibit significantly different levels of false memories than healthy older adults (Balota et al., 1999; Watson et al., 2001). Research with younger adults also indicates those with poor cognitive function will likely perform more poorly on false memory tasks than younger adults with higher cognitive functioning (Gerrie & Garry, 2007; Peters et al., 2006; Peters et al., 2007). Therefore, including individuals with cognitive impairments is problematic. Data from these individuals would likely be outliers and would not be representative of healthy adults.

The MoCA assesses short-term memory, visuospatial skills, executive functions, attention, concentration, working-memory, language, and orientation to time and place. The measure has a maximum score of 30 and a score of 26 or higher is considered in the normal range. However, as discussed below, for older adults a lower cut-off score was employed in the current research. The test and administration instructions are freely accessible from www.mocatest.org. Written permission is required from the MoCA development team when the test is employed in a university research project. Therefore, before the present research commenced written permission to use the MoCA was obtained from the MoCA development team (refer to Appendix E).

The MoCA is reported to be a valid and reliable screening measure. The measure has been found to have excellent sensitivity (the proportion of individuals with cognitive problems which are correctly identified as such) and fair to excellent specificity (the proportion of those without cognitive problems which are correctly identified), when using cut-off scores ranging from 24 to 27, to detect individuals with mild cognitive impairment, Alzheimer's disease or other dementias, and normal older adults (Hoops et al., 2009; Luis, Keegan, & Mullan, 2009; Nasreddine et al., 2005; Smith, Gildeh, & Holmes, 2007). The MoCA is reported to have high test-retest correlation coefficients (Duro, Simões, & Ponciano, 2010; Gill, Freshman, Blender, & Ravina, 2008; Nasreddine et al., 2005; Smith et al., 2007), high internal consistency (Duro et al., 2010; Nasreddine et al., 2005), and convergent validity has been established through moderate to strong correlation coefficients with the MMSE (Gill et al., 2008; Nasreddine et al., 2005). Research has also indicated that the total score of the MoCA yields a reliable estimate of global cognitive ability (Koski, Xie, & Finch, 2009).

As mentioned earlier, the present research employed a cut-off score lower than the score suggested by Nasreddine et al. (2005). Initially, the cut-off score employed was 26. However, after gathering data from a number of older adults it became clear this score was inappropriate. Various studies have demonstrated that a cut-off score of 26 is too high and may over-pathologise individuals who do not have cognitive problems (Luis et al., 2009; Rossetti, Lacritz, Cullum, & Weiner, 2011; Waldron-Perrine & Axelrod, 2012). For instance, Luis et al. (2009) found mean MoCA scores of 25.9 (*SD* = 1.8), 20.5 (*SD* = 2.4), and 15.8 (*SD* = 6.5) for adults with no cognitive problems, mild cognitive impairment, or Alzheimer's disease, respectively. Similar mean scores (25.0, *SD* = 3.1; 22.5, *SD* = 3.5; and 21.0, *SD* = 3.4, respectively) were found by Smith et al. (2007). Luis et al. reported an optimal cut-off score of 23, which resulted in excellent specificity and sensitivity. Additionally, Lee et al. (2008) validated the Korean version of the MoCA and found the optimal cut-off score was 22/23, which provided excellent sensitivity and specificity estimates.

Furthermore, Rossetti et al. (2011) collected normative data for a large sample (*N* = 2,148) of ethnically diverse individuals, aged 18-85 years old, residing in America. The mean MoCA score was 23.36 (*SD* = 3.99). The oldest adults, 70-80 years old, had a mean MoCA score of 21.32 (*SD* = 4.78), and the youngest adults, 18-35 years old, had a mean MoCA score of 25.16 (*SD* = 3.08). Lastly, a study conducted to collect normative data for the MoCA with a New Zealand sample found that adults 65 years and older had a mean score of 23.3 (*SD* = 3.0), whilst adults aged 25 to 34 had a mean score of 27.0 (SD = 1.8; Sothieson, 2010).

The findings discussed above indicate that it is appropriate to reduce the cut-off score of the MoCA, particularly with older adults. In the present research the cut-off score for those aged 75 to 80 was 24. This score represents a trade-off between increasing the probability that individuals without cognitive problems will be identified as such (thus increasing the number of participants that can be included in the final analyses), while also increasing the likelihood that people who do have cognitive problems are included in the research. Based on the discussion above, it seems unlikely that lowering the cut-off score from 26 to 24 will lead to the inclusion of individuals with cognitive problems beyond what would have occurred if the score remained at 26. A score of 24 is also one point above the mean found by Sothieson (2010) using a New Zealand sample. Because there is limited evidence to suggest employing a lower cut-off score for younger adults (e.g., Rossetti et al., 2011; Sothieson, 2010), their cut-off score remained at 26.

Participant questionnaire. The questionnaire was a simple task that included questions regarding the time of day, individuals' ages, if they identify as female or male, and their highest level of education. Time of day was included as a confound check because research has demonstrated that when older and younger adults are tested on false memory tasks at their optimal testing time of day – the morning for older adults and the afternoon/evening for younger adults – their false memory rates are comparable (Intons-Peterson et al., 1999).

Apparatus

To create the DRM recordings the word lists were read by a male and recorded onto a computer. Recording onto computer permitted editing, and thus precise timing of one word every 1.5 seconds. Each list was a separate Windows Media Audio file. Lists were played at an appropriate volume depending on the participant's hearing ability and preferred volume. The recordings were played through two Logitech stereo speakers connected to a Philips GoGear Raga mp3 player which was controlled by the researcher during data collection.

Procedure

Participants were tested individually at either their personal residence or in a room at the Manawatu campus of Massey University (Palmerston North). In all cases it was ensured that the room used to conduct the research was quiet, private, and free from distractions.

Firstly, participants were asked to read an information sheet detailing the nature of the study (Appendix F) and sign a consent form (Appendix G). Following this participants were asked to complete the participant questionnaire (Appendix H). Next, the MoCA was administered (as per the standardised instructions), and took approximately 10 minutes to complete. Following the MoCA, participants began the false memory task. Participants were read (by the researcher) a standard set of instructions appropriate to the experimental condition that they had been assigned to. The instructions included an example of the test (with three words unrelated to the DRM lists) to demonstrate the task. Both sets of instructions can be found in Appendix I. The instructions were repeated if necessary. Once participants indicated they understood the task and were ready to begin, the 120 words were played (through the speakers). At the end of the 120 words participants were handed the memory test to complete in their own time (which took approximately 10-20 minutes).

Finally, because the research involved a low level of concealment, as participants were not specifically told that the study examined susceptibility to false memories, participants were debriefed. Debriefing occurred immediately after data collection with each participant. Participants were first provided with a verbal explanation, in lay terms, of the type of memory ‘mistake’ (false memories in the DRM paradigm) that the research investigated and how they were investigated. Additionally, participants were told this information was somewhat concealed because if they knew exactly what was being examined they might have been able to employ strategies they would not normally use to avoid making the mistakes. Participants were then encouraged to ask any questions they had, or voice any concerns, which were discussed. Participants were also reminded they could contact the researcher or supervisor at any time if they had other concerns or queries.

Lastly, participants were asked not to tell other people about the information discussed during debriefing, or the nature of the task, because these individuals may later participate in the project and having them know such information could affect the results. Participants were then thanked for their time and reimbursed with their choice of a $10 New World, Countdown, or Motor Trade Association voucher.

Chapter Five

Results Part One: Recognition Accuracy

The current research focused on performance on the DRM (Deese/Roediger-McDermott) false memory task across two age groups (16-30 and 75-80 years old) and two test conditions (standard old/new recognition task and source-monitoring task). This chapter first examines and discusses the assumptions that are required to be met when employing multivariate analysis of variance (MANOVA) and analysis of variance (ANOVA). Results from the MoCA and the time of day testing took place are also briefly outlined. These initial analyses are then followed by a discussion of the procedure employed to adjust alpha levels to accommodate the use of multiple significance tests. The chapter then turns to the main analyses, starting with a MANOVA of the four groups' recognition accuracy, followed by a discriminant analysis and separate ANOVAs to examine each dependent variable. Certainty ratings for falsely recognised critical lures and weak lures and correctly recognised presented items are then analysed using ANOVA. Next, response rates for the two age groups in the source-monitoring condition are analysed. Lastly, within-groups pairwise comparisons of recognition accuracy for the different item types (critical lures, weak lures, unrelated items, and presented items) are conducted.

Demographic data for the groups (N = 80) used in most of the following analyses are provided below in Table 5.1. For certainty ratings, some participants did not falsely recognise any critical or weak lures and ratings could not be calculated, resulting in the groups differing from those presented in Table 5.1. These demographic differences will be outlined below with the analyses of certainty ratings (Table 5.3). The demographic data in Table 5.1 shows that all four age and condition groups contained more females than males. For the younger adults the mean age of those in the source-monitoring condition was less than those in the standard condition. Also, the mean MoCA scores appear to differ between older and younger adults, and possibly between the older age groups. These MoCA differences are discussed further below.

Table 5.1

Demographic Data for the Four Age and Test Condition Groups

	Young		Old	
Demographics	Standard task	Source task	Standard task	Source task
n	20	20	20	20
Age range	18-30	16-30	75-80	75-79
Age *M* (*SD*)	23.85 (3.77)	21.80 (4.19)	77.25 (1.55)	77.25 (1.71)
Females	13	12	12	11
Males	7	8	8	9
MoCA *M* (*SD*)	28.05 (1.61)	28.15 (1.42)	27.25 (1.68)	26.35 (1.50)

Initial Analyses

Initial analyses assessed the assumptions required for MANOVA and ANOVA for each of the dependent variables in each of the four age and condition groups. These included examining and testing for outliers, normality, homogeneity of variances and covariance matrices, and correlations between dependent variables (required for MANOVA). The dependent variables assessed were (a) recognition rates for each item type (critical lures, weak lures, unrelated words, and presented items) across the four groups. For the standard condition recognition scores were obtained by calculating the number of *heard* responses. For the source-monitoring condition recognition scores were obtained by calculating and combining the responses *I heard this word* and *I heard this word and thought of it on my own*. To maintain consistency through the present and next chapter recognition scores were transformed into recognition rates; (b) mean certainty ratings for falsely recognised critical lures and weak lures and correctly recognised presented items across the four groups. Certainty ratings for falsely recognised unrelated items were excluded as most of the participants did not recognise any; therefore ratings, could not be calculated for these participants; and (c) source-monitoring response rates of *I heard this word*, *I heard this word and thought of it on my own*, *I did not hear this word but thought of it on my own*, and *I did* *not hear or think of this word*[7] for each item type across the two age groups. Group results from the MoCA and the time testing took place are also briefly outlined below.

Outliers. Based on recommendations by Aguinis et al. (2013), Barnett and Lewis (1994), Field (2009), Pallant (2007), and Tabachnick and Fidell (2013) potential outliers were identified using multiple tools. These tools were box plots, *z*-scores, Mahalanobis distances (which screen for multivariate outliers), leverages, and studentized deleted residuals[8]. Cases were identified as potential outliers if on boxplots they were situated 1.5 times the interquartile range (or more) from the box (Cardinal & Aitken, 2006; Field, 2009), their *z*-score absolute value exceeded 2.5 (Hair, Black, Babin, & Anderson, 2009; Stevens, 2009), their Mahalanobis distance value exceeded a critical value of 10.67 (Barnett & Lewis, 1978), their leverage value exceeded 0.75 (J. Cohen, Cohen, West, & Aiken, 2003; Stevens, 2009), or their studentized deleted residual absolute score exceeded a critical value of 2.13 (Aguinis et al., 2013). Hair et al. (2009) state that cases consistently identified as potential outliers using different tools should be identified as outliers. In the current research, cases identified as potential outliers on a dependent variable for two or more of the various screening tools were defined as outliers.

Using the previously described identification tools and methods, eight cases were identified as outliers across the four item type recognition rates (one for critical lures, two for weak lures, five for unrelated words, two for presented items, and one multivariate outlier). These cases consisted of two younger and three older adults from the standard condition, and two younger and one older adult from the source-monitoring condition. For mean certainty ratings three outliers were identified for falsely recognised critical lure ratings (one younger adult from the source condition, and one older adult from each of the test conditions), none for falsely recognised weak lure ratings, and two for presented item ratings (one older adult from each test condition). For the source-monitoring response options two cases were identified as outliers for critical lure source options (both older adults), two for weak lure source options (both older adults), six for unrelated item source

---

[7] For ease the options will be abbreviated and referred to as *heard only, heard and thought, thought only,* and *neither heard/thought*, respectively.

[8] Although leverages and studentized deleted residuals tend to be regarded as tools to screen for outliers in regression, Aguinis et al. (2013) state their use also applies to ANOVA, because the general linear model is the foundation in both of these analyses.

options (three from each age group), and two for presented item source options (one from each age group). As recommended by Aguinis et al. (2013) and Stevens (2009) two analyses were conducted for each statistical test, one including all participants' data and one excluding outliers from the dependent variable(s) being analysed (pairwise exclusion). Results were then compared for the two analyses. Because findings for the two sets of analyses did not differ in statistical significance, only results from the analyses containing all data are reported below.

Normality. Univariate normality was assessed by examining histograms, Q-Q plots, Kolmogorov-Smirnov normality tests, and skewness and kurtosis values and their respective *z*-scores. The following dependent variables are assessed across age and/or condition groups: recognition rates of each item type; certainty ratings of critical lures, weak lures, and presented items; and source-monitoring response options for each item type.

Recognition proportions. Histograms and Q-Q plots for recognition rates suggested most groups' data for each of the four item types deviated from normality, particularly for critical lures and unrelated items, less so for weak lures and presented words. The Kolmogorov-Smirnov tests revealed recognition rates were not normally distributed for critical lures and unrelated words in all four groups, and presented items for the older adults in the standard recognition condition. Further tests revealed that the following data were significantly skewed: critical lures for older adults in the source condition, weak lures for younger adults in the source condition, and unrelated items for all groups. There were significant leptokurtic distributions for the older group in the standard condition and both age groups in the source-monitoring condition (refer to Appendix J, Table J1 for the Kolmogorov-Smirnov test results, and Table J2 for the skewness and kurtosis *z*-scores).

Certainty ratings. Visual examination of histograms and Q-Q plots for certainty ratings for critical lures, weak lures, and presented items indicated all groups' data deviated from normality. The Kolmogorov-Smirnov tests revealed that the following data were not normally distributed: critical lure ratings for younger adults in the standard condition, and both older adult groups; and presented item ratings for older adults in the standard condition. The following data were significantly skewed: older adults in both conditions critical lure and presented item ratings, and the younger adults in the source condition critical lure ratings. Significant leptokurtic distributions were found for older adults in both conditions for critical lure and presented item ratings (refer to Appendix J, Table J3 for the Kolmogorov-Smirnov test results, and Table J4 for skewness and kurtosis *z*-scores).

Source-monitoring responses. Visual inspection of histograms and normal Q-Q plots for the source-monitoring task response rates across the four item types suggested most if not all data deviated from normality in both the younger and older groups. The Kolmogorov-Smirnov tests indicated that in the younger group, 8, and in the older group, 13, of the 16 dependent variables deviated from normality (refer to Appendix J, Table J5). For the younger adults, 5, and for the older adults, 12, of the 16 dependent variables were significantly skewed. Significant leptokurtic distributions were found for 3 variables for younger adults, and 6 for older adults (for skewness and kurtosis *z*-scores refer to Appendix J, Table J6).

Importantly, Stevens (2009) reviews various research and states deviations from normality have only a small impact on Type I error, and the effect of skewness on univariate (and thus likely multivariate) power is negligible. Platykurtosis can attenuate power, but this is not a concern, as only significant leptokurtic distributions were found in the current data. Consequently, despite a number of variables demonstrating deviations from normality this is unlikely to greatly impact results. With MANOVA, to overcome difficulties with assumption violations alternative test statistics can be considered and the most accurate based on group sizes and assumption violations can be employed. Field (2009) notes that all four of the MANOVA test statistics are robust to violations of normality, and when group sizes are equal the most robust is Pillai's trace. Therefore, the test statistic used in the present research is Pillai's trace.

Homogeneity of variance. Univariate equality of variance was assessed using the Levene's test. Variances for recognition of critical lures, $F(3, 76) = 7.39$, $p < .001$, and presented items, $F(3, 76) = 5.96$, $p = .001$ were not homogenous. Variances for certainty ratings were found to be homogenous. For the source-monitoring condition critical lure responses *thought only*, $F(1, 38) = 24.55$, $p < .001$, and *neither heard/thought*, $F(1, 38) = 4.39$, $p = .04$; and the presented item response *heard only*, $F(1, 38) = 8.28$, $p = .007$, and *neither heard/thought,* $F(1, 38) = 9.42$, $p = .004$ violated the assumption of homogeneity of variance.

Although the equality of variance assumption was not met for some of the variables, Cardinal and Aitken (2006) report that when sample sizes are equal (unless very small) there will be only a minor impact on Type I error rates. Stevens (2009) states that *F* is robust when group sizes are roughly equal (i.e., largest/smallest < 1.5), as they are in the current study. Therefore, for the present research, *F* was assumed to be robust.

In addition to homogeneity of variance (required for ANOVA), when there are more than two groups MANOVA requires equality of covariance matrices. This was assessed via Box's test, and found to be significant ($p < .001$). However, Field (2009) states this test is unstable when sample sizes are equal, and it can be disregarded as Hotelling's and Pillai's trace statistics are assumed to be robust in the face of equality of covariance matrices violations (another reason why Pillai's trace is employed in the current research).

Correlations between dependent variables. Prior to conducting a MANOVA it is important to check that the dependent variables are correlated moderately (Meyers, Gamst, & Guarino, 2013) or highly (Cole, Maxwell, Arvey, & Salas, 1994; Stevens, 1980) with each other. This can be done by examining the correlations between dependent variables for each group and by using Bartlett's test of sphericity when there are more than two groups. For Bartlett's test a significant result indicates adequate correlations exist among the dependent variables to advance with a MANOVA (Hair et al., 2009; Meyers et al., 2013).

Recognition proportions. Significant results were found for Bartlett's test for recognition proportions of the four item types for all four age and condition groups: younger adults in the standard condition, $\chi^2(6) = 40.01$, $p < .001$; older adults in the standard condition, $\chi^2(6) = 15.12$, $p = .02$; younger adults in the source-monitoring condition, $\chi^2(6) = 45.54$, $p < .001$; and older adults in the source-monitoring condition, $\chi^2(6) = 37.92$, $p < .001$. For each age group most of the Pearson's correlations between the dependent variables ranged from moderate to high (refer to Appendix J, Table J7). Overall, these results indicate appropriate use of a MANOVA.

Source-monitoring responses. Pearson's correlations were examined for the source-monitoring responses for each item type across the two age groups. For critical lure, weak lure, and unrelated responses correlations for both age groups ranged from being negligible to high, with approximately half of the correlations being moderate or high for each dependent variable and age group. For presented item responses correlations ranged from moderate to high for younger adults, and negligible to high for older adults with half of the correlations being moderate or high (these correlations can be viewed in Appendix J, Table J8). Although not all of the Pearson's correlations for the source-monitoring responses were deemed moderate or high, the majority of correlations were, suggesting it is appropriate to use MANOVAs.

The MoCA. Although the MoCA was only used as a screening tool to establish participant inclusion/exclusion, a two-way between groups ANOVA was conducted to examine any age and condition group differences of those included in the present study. Before conducting the ANOVA the necessary assumptions were checked, and all were satisfied (as this is not a main analysis these assumptions will not be discussed further). The results found older adults (M = 26.80, SD = 1.64) performed more poorly than younger adults (M = 28.10, SD = 1.50), $F(1, 76) = 13.98$, $p < .001$, $\eta^2 = .15$. There were no differences between the test conditions, $F(1, 76) = 1.32$, $p = .254$, $\eta^2 = .01$, or an interaction effect, $F(1, 76) = 2.07$, $p < .155$, $\eta^2 = .02$. These results are unsurprising, as one would expect to find age differences in performance for a cognitive measure, particularly when comparing adults of such different ages (i.e., 16-30 compared to 75-80 years old). Such findings are in keeping with those of Rossetti et al. (2011), in their attempt to establish normative data, discussed earlier.

Time testing took place. As outlined previously, the time of day that testing took place was recorded for each participant as a confound check, because testing participants at their optimal testing time can affect age-related false memory findings (Intons-Peterson et al., 1999). Visual inspection of histograms demonstrated that a similar number of participants across the four age and condition groups were tested during the morning (8:00-11:00 a.m.), midday (11:00 a.m.-1:00 p.m.), and the afternoon (1:00-5:00 p.m.).

Adjusted Alpha Level

To account for the inflated Type I error rate when multiple tests are conducted on the same set of data the Holm (1979) procedure was employed. The Holm procedure is a simple step-down, sequentially rejective, procedure in which a family of p values are ranked smallest to largest and, starting with the smallest value, each is tested with the adjusted alpha (α’) level:

$$\alpha' = \frac{\alpha}{n - i + 1} \tag{5.1}$$

Where α is the omnibus criterion of significance, *n* is the number of hypothesis tests in the family, and *i* is the rank position of the significance value. If the first significance value is found to be significant the second value is tested, and so on until a value is found not significant. The Holm procedure can also be conducted by sequentially adjusting each significance value (p'), whilst the alpha level is kept constant, using:

$$p' = (n - i + 1)\, p \tag{5.2}$$

For ease the current research will present unadjusted and adjusted significance values, instead of the adjusted alpha levels, unless $p < .001$, in which case adjusted alpha levels will be provided as adjusted significance values cannot be accurately computed.

The Holm (1979) procedure was employed because it does not require the test statistics involved to meet any assumptions (e.g., be independent or have parametric distributions), making it applicable to use in a range of situations (Farcomeni, 2007; Ge, Sealfon, Tseng, & Speed, 2007; Holm, 1979). It is easy to apply and is more powerful than other methods (e.g., the Bonferroni method), while still maintaining strong control over Type I error (Ge et al., 2007; Holland & Copenhaver, 1987). Additionally, the Holm procedure can be used to control overall Type I error across the various tests conducted within a family, instead of setting a single overly conservative error rate for each test (Keselman, Miller, & Holland, 2011). This allows the multiple MANOVAs conducted in the current research to be included in a family, and have Type I error controlled (as recommended by Tabachnick & Fidell, 2013).

With multiple hypotheses testing a family can be defined in numerous ways (An, Xu, & Brooks, 2013; Farcomeni, 2007; Shaffer, 1995). Generally, a family is defined as a collection of theoretically-related hypothesis tests (Keselman et al., 2011). In the current chapter there are nine families of tests where the Holm procedure is used. The first is the ANOVAs that follow the significant MANOVA examining memory differences across the age and test conditions, with the four item types' recognition rates as dependent variables. This family consists of 12 significance tests, and was chosen because performing alpha adjustments on ANOVAs following a significant MANOVA is recommended by numerous authors (e.g., Field, 2009; Stevens, 2009; Tabachnick & Fidell, 2013). The MANOVA is excluded from the adjustment because it is the main test conducted in the current research.

It is treated as a single family and already has Type I error controlled (Stevens, 2009; Tabachnick & Fidell, 2013). The discriminant analysis is excluded from an alpha level adjustment, because both the *F*-test significance value and the amount of variance a function explains are used to assess the importance of a function (Spicer, 2005).

The second, third, and fourth families are the two-way ANOVAs conducted on the certainty ratings of critical lures, weak lures, and presented items. Each of these families consists of three significance tests, and were chosen as each ANOVA examines group differences for a specific certainty rating. The fifth family consists of the four MANOVAs and subsequent ANOVA conducted to examine differences between the two age groups in the source-monitoring condition on their response options (*heard only, heard and thought, thought only,* and *neither heard/thought*) for each item type. This family consists of eight significance tests, and was chosen because Tabachnick and Fidell (2013) state that when multiple MANOVAs are conducted Type I error needs to be controlled through an alpha level adjustment. As stated previously, ANOVAs that follow a significant MANOVA require Type I error control (e.g., Field, 2009; Stevens, 2009; Tabachnick & Fidell, 2013). Finally, the last four families (each consisting of six significance tests) are the within-subjects pairwise comparisons of the four item types' recognition rates.

Memory Mistakes across Age and Recognition Test Conditions

Bray and Maxwell (1985) assert that a MANOVA is suitable when research examines a set of dependent variables that represent an underlying construct, or which of a set of dependent variables contribute the most to separating groups. Both of these ideas are important in the present research, and MANOVAs have been used in other research examining false memories (e.g., Brueckner & Moritz, 2009; Heaps & Nash, 2001; Murphy, West, Armilio, Craik, & Stuss, 2007). Therefore, the following analysis uses a two-way between-groups MANOVA to investigate memory differences across the two age groups and two test conditions, with the four item type (critical lures, weak lures, unrelated items, and presented items) recognition rates as dependent variables. Table 5.2 below provides the means and standard deviations for recognition rates across the four groups. As recommended by Bray and Maxwell (1982) and Field (2009), to achieve an in depth understanding of the data the significant MANOVA was followed by a discriminant analysis and separate two-way between-groups ANOVAs.

Table 5.2

Mean Recognition Proportions (and SD) for the Four Age and Condition Groups

	Young		Old	
Item type	Standard	Source	Standard	Source
Critical lure	.76 (.17)	.49 (.31)	.83 (.17)	.75 (.20)
Weak lure	.33 (.23)	.25 (.15)	.43 (.24)	.24 (.19)
Unrelated	.12 (.14)	.06 (.06)	.14 (.19)	.10 (.13)
Presented	.69 (.09)	.71 (.10)	.69 (.17)	.59 (.20)

MANOVA. Results of the MANOVA, using Pillai's trace, found that recognition performance differed between age groups, $F(4, 73) = 6.22$, $p < .001$; $V = .25$, $\omega^2_{multivariate} = .22$, and test conditions, $F(4, 73) = 4.40$, $p = .003$; V = .19, $\omega^2_{multivariate} = .16$. Based on the figures presented in Table 5.2 these significant findings indicate younger adults performed better than older adults, and those in the source condition performed better than those in the standard condition. Additionally, age and test conditions were found to interact in their effect on recognition performance, $F(4, 73) = 3.73$, $p = .008$; $V = .17$, $\omega^2_{multivariate} = .14$. Further analyses are required to identify the nature of this interaction.

Discriminant Analysis. As four age and condition groups were involved, the follow up discriminant analysis demonstrated three discriminant functions. All three functions had a joint χ^2 (12) = 48.24, $p < .001$, $\Lambda = .53$, indicating that the three functions accounted for 100(1 - .53) = 47% of the variance in the four age and condition groups. The test of functions two and three was significant, χ^2 (6) = 12.87, $p = .05$, $\Lambda = .84$, and accounted for 100(1 - .84) = 16% of the variance in the four age and condition groups. The third function did not achieve statistical significance, χ^2 (2) = 1.55, $p = .46$, $\Lambda = .98$, and only accounted for 100(1 - .98) = 2% of the variance in the four groups. Functions one and two together accounted for 97% of the explained variance. Function one explained 76.6% of the explained variance, canonical R^2 = .38, function two explained 20.7%, canonical R^2 = .14, function three explained only 2.7%, canonical R^2 = .02.

The structure (loading) matrix[9] of correlations between outcomes and discriminant functions revealed that for the first function critical lure recognition was the best variable for distinguishing groups, as it was the only variable to load highly (*r* = .77) on this function. An examination of group centroids found that the first function separated the younger adults in the source-monitoring condition from the three other groups (see Figure 5.1 below, and the group means and standard deviations provided above in Table 5.2). Put simply, these findings indicate that function one, where critical lure recognition is the main contributor, differentiates the younger adults in the source-monitoring condition from the other three groups.

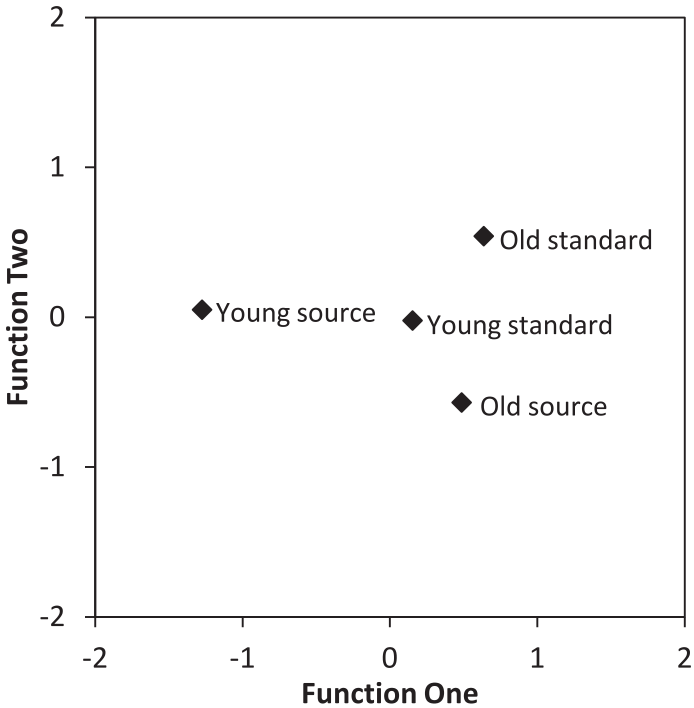


Figure 5.1. Canonical discriminant functions for each age and condition group on the two functions.

For the second function weak lure recognition loaded the most highly (*r* = .83), followed by presented item recognition (*r* = .61). Examining the group centroids showed that the second function discriminated the younger adults in the standard recognition condition *and* the older adults in the source-monitoring condition *from* the younger adults in the source-monitoring condition *and* older adults in the standard recognition condition (see

[9] As recommended by Tabachnick and Fidell (2013) only factor loadings greater than .05 were reported and interpreted for the two functions.

Figure 5.1, and the group means and standard deviations provided in Table 5.2). In other words, function two, where weak lure and presented item recognition contribute the most, clusters two groups (younger adults in the standard condition *and* the older adults in the source-monitoring condition) and differentiates them from the other two clustered groups (younger adults in the source-monitoring condition *and* older adults in the standard condition) in the analysis.

Recognition Accuracy of Each Item Type. The separate two-way ANOVAs employed to compare the dependent variables separately (using the Holm procedure to calculate α' and *p′* values) demonstrated that older adults (M = .79, SD = .19) falsely recognised more critical lures than younger adults (M = .63, SD = .28), $F(1, 76) = 10.96$, $p' = .01$, $p = .001$, $\eta^2 = .11$ (refer to Figure 5.2 below). Those in the source-monitoring condition (M = .62, SD = .29) falsely recognised fewer critical lures than those in the standard test condition (M = .79, SD = .17), $F(1, 76) = 12.71$, $p' = .01$, $p = .001$, $\eta^2 = .12$ (refer to Figure 5.3 below). The lack of an interaction effect indicated that the change in test conditions impacted older and younger adults' critical lure recognition rates equally, $F(1, 76) = 4.15$, $p' = .41$, $p = .05$, $\eta^2 = .04$.

ANOVA results examining weak lure recognition revealed no age group differences, $F(1, 76) = 0.96$, $p' = .93$, $p = .33$, $\eta^2 = .01$ (Figure 5.2 below). By contrast, irrespective of age, those in the source-monitoring condition (M = .24, SD = .17) falsely recognised fewer weak lures than those in the standard recognition condition (M = .38, SD = .24), $F(1, 76) = 8.66$, $p' = .04$, $p = .004$, $\eta^2 = .10$ (refer to Figure 5.3 below). The lack of an interaction effect indicated that the change in test conditions impacted older and younger adults equally (i.e., the interaction effect was not significant), $F(1, 76) = 1.81$, $p' = .91$, $p = .18$, $\eta^2 = .01$.

For recognition of unrelated items no differences were found between age groups, $F(1, 76) = 1.04$, $p' = .93$, $p = .31$, $\eta^2 = .01$, or test conditions, $F(1, 76) = 2.66$, $p' = .76$, $p = .17$, $\eta^2 = .03$ (see Figures 5.2 and 5.3). The interaction was also non-significant, $F(1, 76) = 0.17$, $p' = .93$, $p = .69$, $\eta^2 < .01$.

Lastly, for recognition of presented items no differences were found between age groups, $F(1, 76) = 2.86$, $p' = .76$, $p = .10$, $\eta^2 = .03$, or test conditions, $F(1, 76) = 1.56$, $p' = .91$, $p = .22$, $\eta^2 = .02$ (see Figures 5.2 and 5.3). The interaction was also non-significant, $F(1, 76) = 2.86$, $p' = .76$, $p = .1$, $\eta^2 = .03$.

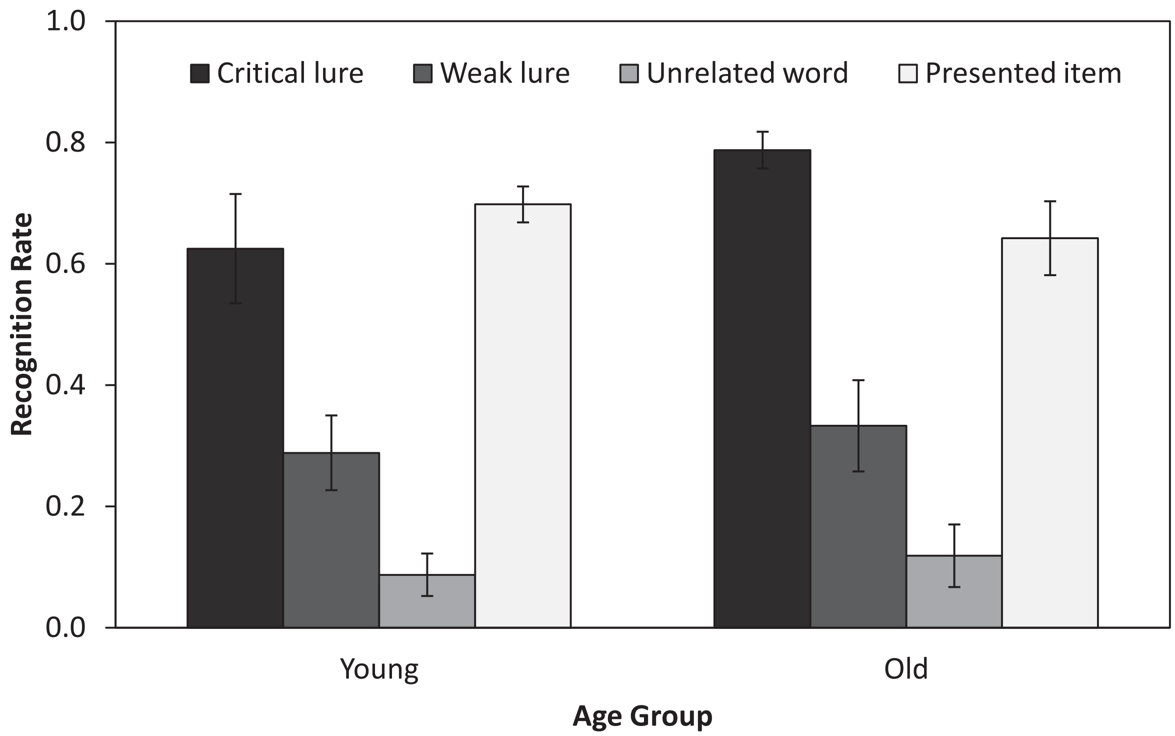


Figure 5.2. Mean recognition rates for the four item types across the two age groups. Vertical lines represent 95% CIs

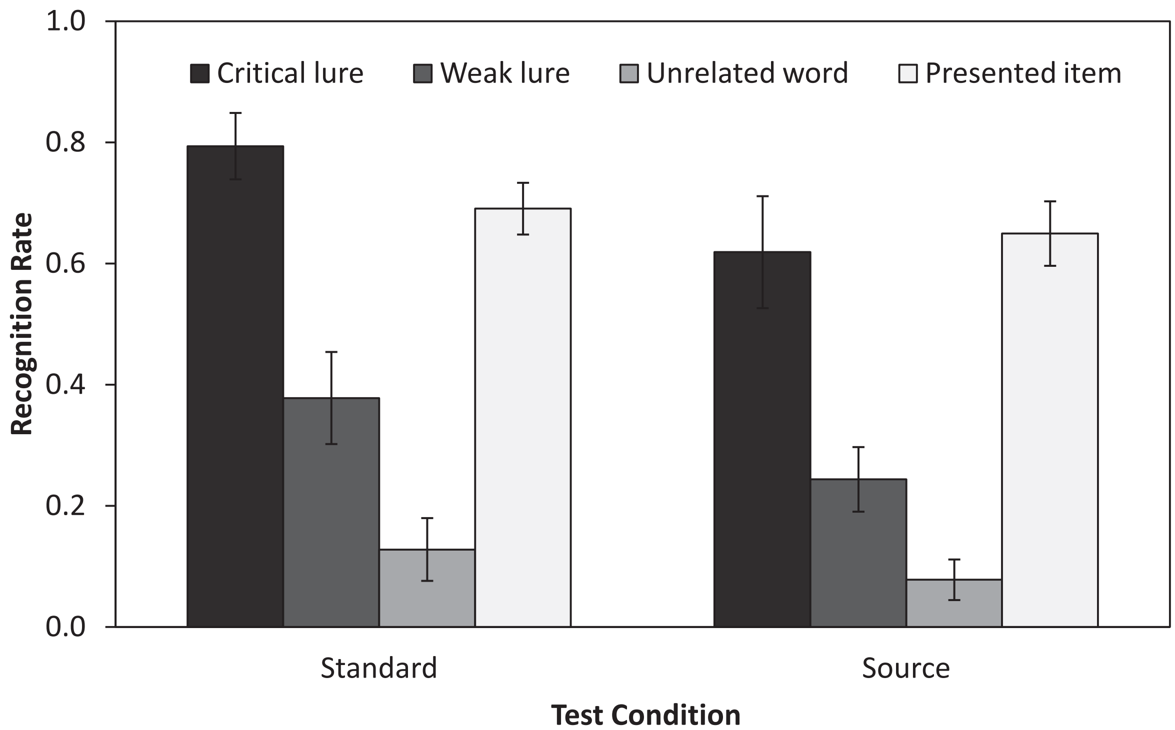


Figure 5.3. Mean recognition rates for the four item types across the two test conditions. Vertical lines represent 95% CIs.

Certainty Ratings

To recap, certainty ratings were obtained by employing a 6-point rating scale for how certain a word was heard or not. The six ratings and corresponding values were: 1 = *very certain heard*, 2 = *mostly certain heard*, 3 = *fairly certain heard*, 4 = *fairly certain not heard*, 5 = *mostly certain not heard*, and 6 = *very certain not heard*. Table 5.3 below provides the mean certainty ratings for each of the four groups for falsely recognised critical and weak lures, and correctly recognised presented items.

Table 5.3

Mean Certainty Ratings (and SD) for the Four Age and Condition Groups for Item Types

	Young		Old	
Item type	Standard	Source	Standard	Source
Critical lure	1.57 (.44)	1.60 (.51)[a]	1.30 (.36)	1.37 (.44)
Weak lure	2.43 (.38)[b]	2.26 (.50)	1.73 (.58)	1.77 (.63)[c]
Presented item	1.58 (.24)	1.46 (.16)	1.34 (.29)	1.40 (.36)

Note. Groups identified with superscript letters differ demographically from those presented in Table 5.1 above. The differences are:

[a] *n* = 18, age *M* = 21.94 (*SD* = 4.37), males = 6. [b] *n* = 19, age *M* = 23.74 (*SD* = 3.84), males = 6. [c] *n* = 18, age *M* = 77.50 (*SD* = 1.62), males = 7.

The two-way ANOVA conducted on certainty rating means for falsely recognised critical lures demonstrated that, younger adults (*M* = 1.58, *SD* = 0.47) were *less* certain than older adults (*M* = 1.34, *SD* = 0.40) that critical lures *had* been presented, $F(1, 74) = 6.14$, $p = .05$, $\eta^2 = .08$ (refer to Figure 5.4 below). No difference was found between the test conditions, $F(1, 74) = 0.23$, $p' = 1$, $p = .63$, $\eta^2 < .01$. The interaction effect was not significant, $F(1, 74) = 0.05$, $p' = 1$, $p = .83$, $\eta^2 < .01$, indicating the different test conditions impacted older and younger adults equally.

The two-way ANOVA conducted on certainty rating means for falsely recognised weak lures found younger adults (*M* = 2.34, *SD* = 0.44) were *less* certain than older adults (*M* = 1.75, *SD* = 0.59) that weak lures *had* been presented, $F(1, 73) = 24.35$, $p < .001$, $\alpha' = .02$, η^2

= .25 (see Figure 5.4 below). There was no difference between the test conditions, $F(1, 73)$ = 0.29, p' = .73, p = .59, $\eta^2 < .01$. The interaction was non-significant, $F(1, 73)$ = 0.82, p' = .73, p = .37, η^2 = .01.

Finally, for certainty ratings of correctly recognised presented items older adults (M = 1.36, SD = 0.32) were found to be *more* certain than younger adults (M = 1.52, SD = 0.21) that presented items *had* been heard, $F(1, 76)$ = 6.38, p' = .04, p = .01, η^2 = .08 (refer to Figure 5.4 below). No difference was found between the test conditions, $F(1, 76)$ = 0.30, p' and p = .59, $\eta^2 < .01$. The interaction effect was not significant, $F(1, 76)$ = 2.35, p' = .26, p = .13, η^2 = .03.

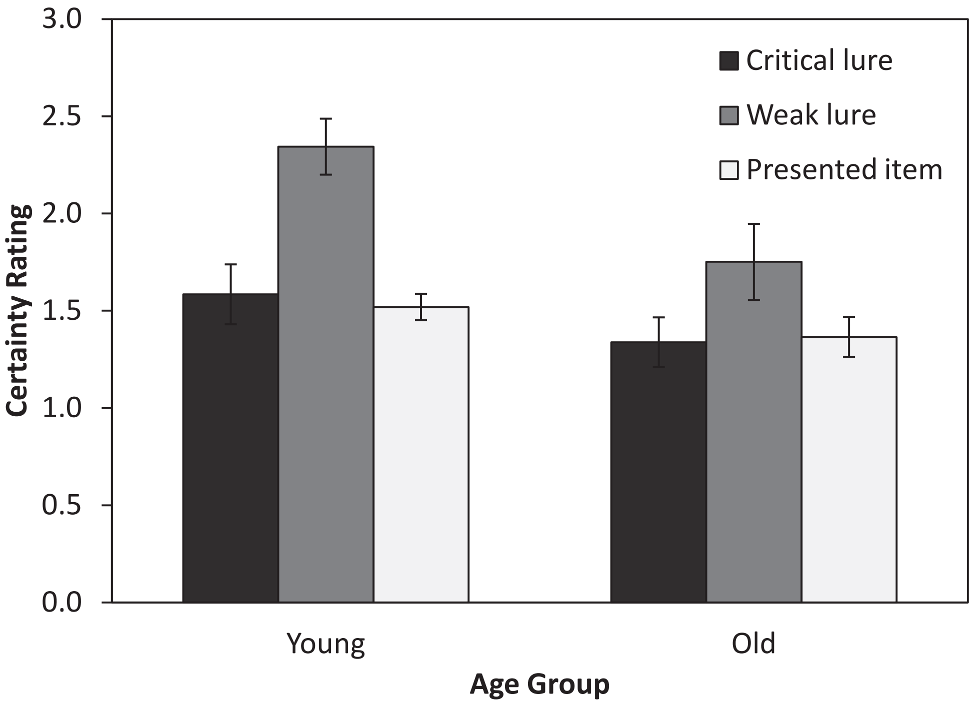


Figure 5.4. Mean certainty ratings for three item types for the two age groups. Vertical lines show 95% CIs.

Memory Mistakes in the Source-Monitoring Condition

The results above are those for the age and condition groups. However, the source-monitoring condition can be examined at a more fine-grained level of analysis than the standard condition. To examine age group differences for the source-monitoring responses (*heard only, heard and thought, thought only,* and *neither heard/thought*) a MANOVA was employed for each item type (critical lures, weak lures, unrelated words, and presented items). Table 5.4 provides group means and standard deviations for these analyses.

Table 5.4

Source-Monitoring Condition Response Rates for the Two Age Groups

	Young	Old
Variable	*M* (*SD*)	*M* (*SD*)
Critical lure response options		
Heard only	.41 (.27)	.71 (.23)
Heard and thought	.08 (.09)	.04 (.12)
Thought only	.39 (.32)	.09 (.11)
Neither heard/thought	.12 (.12)	.16 (.21)
Weak lure response options		
Heard only	.17 (.11)	.19 (.19)
Heard and thought	.08 (.10)	.04 (.08)
Thought only	.18 (.13)	.10 (.14)
Neither heard/thought	.57 (.18)	.66 (.24)
Unrelated item response options		
Heard only	.04 (.07)	.09 (.12)
Heard and thought	.02(.03)	.01 (.03)
Thought only	.06 (.10)	.07 (.15)
Neither heard/thought	.89 (.12)	.83 (.20)
Presented item response options		
Heard only	.62 (.11)	.56 (.20)
Heard and thought	.08 (.08)	.04 (.05)
Thought only	.08 (.05)	.05 (.07)
Neither heard/thought	.22 (.10)	.36 (.22)

Note. These rates are calculated from the number of falsely recognised critical lures, weak lures, and unrelated items, and correctly recognised presented words.

Multivariate analyses were employed instead of repeated-measures ANOVAs because the assumption of sphericity was violated for all four of the repeated measures variables (refer to Appendix J, Table J9). Stevens (2009) states that when sphericity is violated and sample sizes are greater than *the number of repeated observations + 10* (such is the case in the current research), then multivariate analyses are more powerful than univariate procedures. One of the four MANOVAs found significant results, and this was followed up with separate one-way ANOVAs (as indicated by the procedures employed by O'Brien & Kaiser, 1985; Stevens, 2009; Vasey & Thayer, 1987).

The first MANOVA (and follow-up ANOVAs) analysed the age group differences for the critical lure responses. Using Pillai's trace, a significant difference between the two age groups was found, $F(3, 36) = 6.90$, $p = .001$, $p' = .006$, $V = .37$, $\omega^2_{\text{multivariate}} = .31$. Results of the follow-up ANOVAs showed older adults were more likely to respond that they heard critical lures, and less likely to respond that they only thought of critical lures, than their younger counterparts ($F[1, 38] = 14.55$, $p < .001$, $\alpha' = .006$, $\eta^2 = .28$ and $F[1, 38] = 16.44$, $p < .001$, $\alpha' = .006$, $\eta^2 = .30$, respectively). These results are shown graphically in Figure 5.5. Both results represent poorer performance by the older adults in identifying the correct source of critical lures.

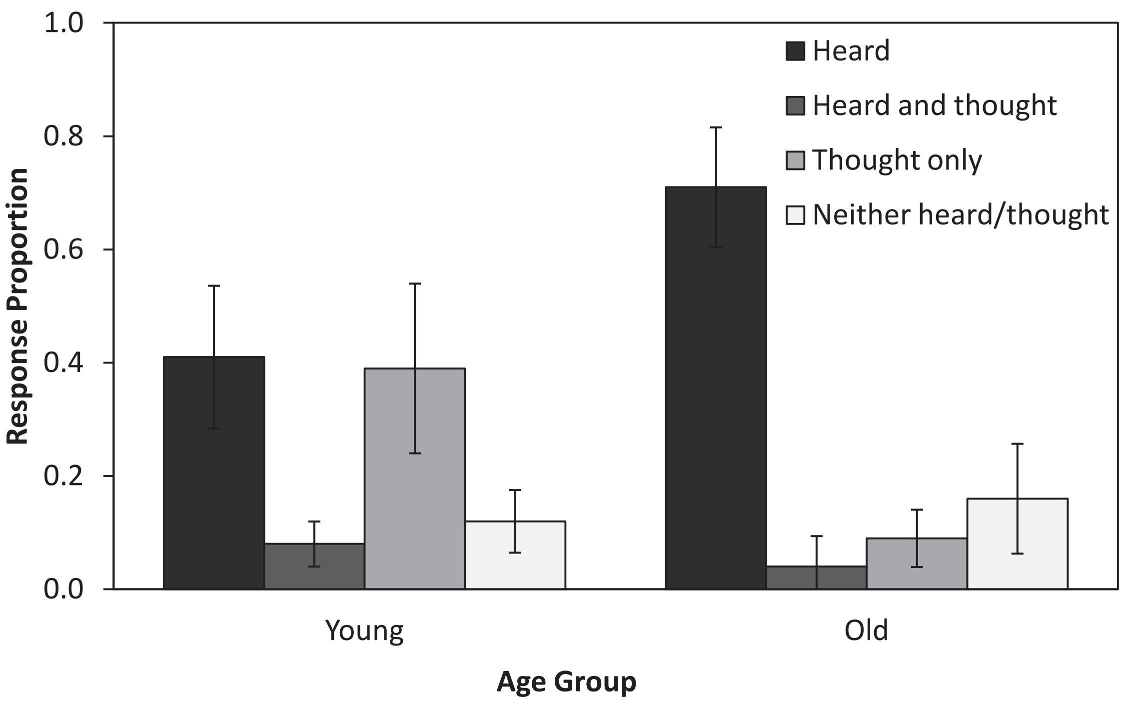


Figure 5.5. Critical lure mean response option rates for younger and older adults in the source-monitoring task condition. Vertical lines show 95% CIs.

The MANOVAs that analysed the source-monitoring responses for weak lures, unrelated items, and presented items (refer to Table 5.5 for group means and standard deviations) found no age group differences. Using Pillai's trace: $F(3, 36) = 1.28$, $p = .30$, $p' = .10$, $V = .10$, $\omega^2_{multivariate} = .02$; $F(3, 36) = 1.10$, $p = .36$, $p' = .10$, $V = .08$, $\omega^2_{multivariate} = .01$; and $F(3, 36) = 2.94$, $p = .05$, $p' = .23$, $V = .10$, $\omega^2_{multivariate} = .13$, respectively.

Pairwise Comparisons of Recognition Rates

The above analyses have focused on between-group differences. By contrast, the following analyses focus on within-group (the four age and condition groups) differences for recognition rates of the four item types. Similar comparisons have been utilised in previous DRM research (e.g., Intons-Peterson et al., 1999; Roediger & McDermott, 1995; Tun et al., 1998) to examine if false recognition of critical lures occurs at the same rate as correct recognition. The following analyses include six pairwise comparisons between the four different item types for each of the four groups. Recognition means for the four groups were provided previously in Table 5.2.

Results from the pairwise comparisons demonstrated that within each age and condition combination the four item types' recognition rates were significantly different from one another (p' and $p < .05$ for all 24 comparisons). As can be seen in Figure 5.6 below, the younger adults in the standard condition and both of the older groups' results followed the same pattern. Critical lure recognition was the highest, followed by presented items, then weak lures, and lastly unrelated items (critical lures > presented words > weak lures > unrelated words). These results indicate that for these three groups the false memory effect was strong. However, the effect was not as strong for the younger adults in the source-monitoring condition, whose results followed a different pattern. Presented items had the highest recognition rate, followed by critical lures, then weak lures, and lastly unrelated words (presented words > critical lures > weak lures > unrelated words).

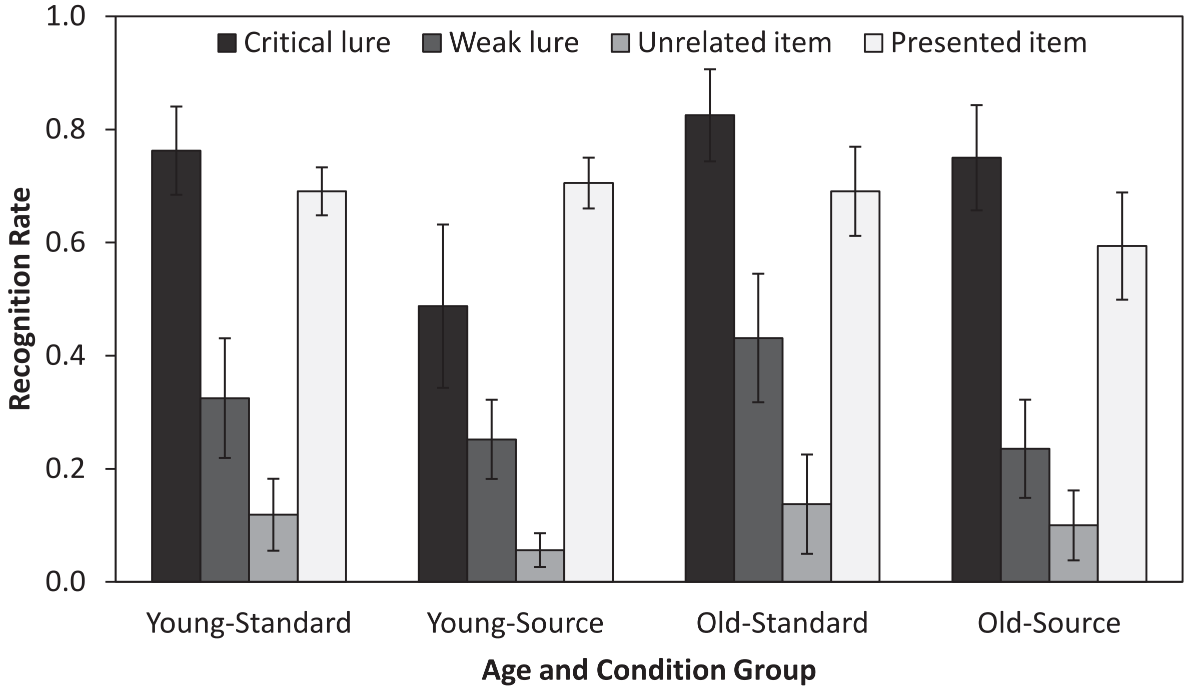


Figure 5.6. Mean recognition rates of the four item types across the four age and condition groups. Vertical lines represent 95% CIs.

Summary

Data from two age groups (16-30 and 75-80 years old) and two test conditions (standard old/new recognition task and source-monitoring task) for a DRM false memory task were examined using multivariate and univariate analyses of variance. The main MANOVA indicated both age and condition group differences on the combined dependent variables (critical lure, weak lure, unrelated item, and presented item recognition). The subsequent discriminant analysis revealed two significant functions. The first function showed that critical lure false recognition was best at differentiating groups, and separated the younger source-monitoring group from the three other groups. The second function found that weak lure false recognition and presented item recognition (as a combination) differentiated the younger standard test group *and* the older source-monitoring group *from* the other two groups (the younger source-monitoring and older standard test condition groups).

The ANOVAs that followed the MANOVA demonstrated that older adults falsely recognised more critical lures than younger adults, whilst no age-related differences were found for weak lures, unrelated words, or presented items. The ANOVAs also showed that those in the standard test condition falsely recognised more critical lures and weak lures than those in the source-monitoring condition. By contrast no differences between the test conditions were found for recognition of unrelated or presented items. Lastly, no significant interactions were found, indicating that the different test conditions impacted older and younger adults' recognition of the four types of test items similarly.

ANOVAs were employed to examine certainty ratings for falsely recognised critical lures and weak lures, and correctly recognised presented items. Age-related differences were found for each of the three item types' ratings, with younger adults being less certain than older adults that they had been presented. No differences between test conditions were found. No significant interactions were found, suggesting the different test conditions influenced participants' certainty similarly across age groups.

Examination of the source-monitoring responses found older adults more often responded that critical lures were *heard* compared to younger adults, while younger adults more often responded that critical lures were *thought only* compared to older adults. Both of these results indicate older adults perform more poorly at identifying the correct source of critical lures compared to younger adults. No age-related differences were found for source responses of weak lures, unrelated items, or presented items. Lastly, a within-groups examination of recognition rates demonstrated that for each of the four age and condition groups recognition rates of the four item types differed significantly from one-another. The pattern of recognition results also differed between younger adults in the source-monitoring group (presented words > critical lures > weak lures > unrelated words) and the three other groups (critical lures > presented words > weak lures > unrelated words).

Chapter Six

Results Part Two: Signal Detection Analyses

The previous chapter examined age and condition group differences in recognition performance on the DRM (Deese/Roediger-McDermott) task by employing multivariate and univariate analyses of variance, which are the types of analyses generally utilised in DRM research. The present chapter goes beyond these more common analyses by utilizing signal detection measures of sensitivity and response bias. Noteworthy, in the current chapter, in line with signal detection theory, different terms (than those in the previous chapter) will be used to refer to participants' recognition performance. That is, results previously referred to as recognition rates will now be called hit rates and false alarm rates. Hit rates are the proportion of presented items correctly identified as old/heard, and false alarm rates are the proportion of non-presented items incorrectly identified as old/heard. Furthermore, although there is only a single measure of hit rate, the following four false alarm rates are utilised and different sensitivity and criterion indices are calculated for each: overall false alarm rate, critical lure false alarm rates, weak lure false alarm rates, and unrelated item false alarm rates. For reference, Table 6.1 below provides a glossary of these terms and their corresponding signal detection indices.

This chapter first provides a brief overview of the signal detection measures employed. Then, because these measures were compared (across the four age and condition groups) using ANOVAs, the required assumptions are examined and discussed. Like the previous chapter, the Holm (1979) procedure is employed to adjust alpha levels to accommodate the use of multiple significance tests. Therefore, there is a section that outlines the families of significance tests which the Holm procedure is applied to. Results are then presented for sensitivity and bias using the overall false alarm rate, and then for the other false alarm types.

Table 6.1

Glossary of Signal Detection Result Terms

Hit rate	The proportion of presented items correctly identified as old/heard
Overall false alarm rate	The combined total proportion of critical lures, weak lures, and unrelated words incorrectly identified as old/heard
Overall sensitivity/*d’*	Sensitivity (*d’*) calculated from the overall false alarm rate and the hit rate
Overall response bias/*c*	Response bias (*c*) calculated from the overall false alarm rate and the hit rate
Critical lure false alarm rate	Proportion of critical lures incorrectly identified as old/heard
Critical lure sensitivity/*d’*	Sensitivity (*d’*) calculated from the critical lure false alarm rate and the hit rate.
Critical lure response bias/*c*	Response bias (*c*) calculated from the critical lure false alarm rate and the hit rate
Weak lure false alarms rate	Proportion of weak lures incorrectly identified as old/heard
Weak lure sensitivity/*d’*	Sensitivity (*d’*) calculated from the weak lure false alarm rate and the hit rate
Weak lure response bias/*c*	Response bias (*c*) calculated from the weak lure false alarm rate and the hit rate
Unrelated item false alarm rate	Proportion of unrelated words/items incorrectly identified as old/heard
Unrelated item sensitivity/*d’*	Sensitivity (*d’*) calculated from the unrelated false alarm rate and the hit rate.
Unrelated item response bias/*c*	Response bias (*c*) calculated from the unrelated false alarm rate and the hit rate

Sensitivity and Response Bias

As discussed in Chapter 3, sensitivity is the ability to discriminate signal from noise (Green & Swets, 1966). In the current research sensitivity is the ability to discriminate presented words (old items) from critical lures, weak lures and/or unrelated words (new items) at test. The sensitivity index employed is *d'*, defined as (Macmillan & Creelman, 2005):

$$d' = z(\mathrm{H}) - z(\mathrm{F}), \tag{6.1}$$

where *z* is the inverse of the normal distribution (H and F refer to the hit rate and false alarm rate, respectively), and measures (using standard deviations) the distance between the signal mean and the noise mean. When *d'* is at (or near) zero it indicates an inability to differentiate between signal and noise, whereas higher values indicate better sensitivity (Green & Swets, 1966; McNicol, 1972). Negative values of *d'* can arise when false alarm rates exceed hit rates (Stanislaw & Todorov, 1999).

Response bias is the tendency to favour one response over another (either old/heard or new/not heard), and is based on one's criterion index (Macmillan & Creelman, 2005). As recommended by Macmillan and Creelman, and Snodgrass and Corwin (1988), the criterion index employed presently is *c*, and is the distance from the neutral criterion point of zero (which indicates no response bias). A positive value of *c* indicates a bias to respond no/new, while a negative value indicates a bias to respond yes/old (Macmillan & Creelman, 2005) The criterion value, *c*, is defined as:

$$c = -\frac{z(\mathrm{H}) + z(\mathrm{F})}{2} \tag{6.2}$$

Difficulties can arise when hit rates or false-alarm rates are zero or one, because the corresponding *z*-scores are infinite, which results in undefined *d'* and *c* scores. To avoid infinite *z*-scores an adjustment needs to be made to the hit and false alarm rates (Macmillan & Creelman, 2005). As recommended by Stanislaw and Todorov (1999), the present research employed the loglinear approach, in which, 0.5 is added to each participant's number of hits and false alarms, and 1 is added to both the total number of signal trials and the total number of noise trials. Hit and false alarm rates are then calculated using these new values.

In the current analyses this transformation was utilized for hits and critical lure, weak lure, and unrelated false alarm rates when calculating *d′* and *c* for the item types separately, but not for calculating *d′* and *c* based on the overall false alarm rate (as there were not hit rates or false alarm rates of 0 or 1 when the false alarm types were combined).

Because *d′* and *c* are based on the assumption that the underlying signal and noise distributions are normal with equal variances, McNicol (1972) recommends calculating alternative indices to accommodate the possibility that such assumptions are not met with the data. Thus, for sensitivity, *A′* and A_g were calculated (Macmillan & Creelman, 2005). The index *A′* is the area under a one-point ROC curve, and is calculated using the following equation from Snodgrass and Corwin (1988):

$$A' = \begin{cases} 0.5 + \dfrac{(H - F)(1 + H - F)}{4H(1 - H)} & \text{When } H \geq F \\ 0.5 - \dfrac{(F - H)(1 + F - H)}{4F(1 - H)} & \text{When } H < F \end{cases} \qquad (6.3)$$

A_g (also referred to as *P(A)*; McNicol, 1972) is the area under a multi-point ROC curve, and is calculated as:

$$A_g = 0.5 \sum (F_{i+1} - F_i)(H_{i+1} + H_i), \qquad (6.4)$$

where (H_i, F_i) refers to the consecutive points on the ROC curve, starting from the point at the lower left corner (at which point H = F = 0.00). Appendix K, Figures K1-K16, provides ROC curves for each group created from hit and overall false alarm rates, as well as for each of the separate false alarm types. Values of *A′* and A_g generally range from .5 (indicating an inability to discriminate signal from noise) to 1 (indicating perfect performance), but, values less than .5 can arise when false alarm rates are higher than hit rates (Stanislaw & Todorov, 1999).

The alternative bias measure employed was a version of *c*, named c_2, which is similar to *c* but takes into account the slope of the z-transformed ROC (refer to Appendix L, Figures L1-L16 for *z*ROC curves). This alternative criterion is calculated as:

$$c_2 = -(s / [1 + s])(z[\mathrm{H}] + z[\mathrm{F}]), \quad (6.5)$$

the slope, s, is obtained by estimating the line of best fit to the zROC (Green & Swets, 1966; Macmillan & Creelman, 2005). The indices A_g and c_2 were calculated on a group-by-group basis, not individually as was done for d', A', and c. This is because A_g and c_2 require ROC curve data; that is, signal and noise responses for each of the six rating categories. Numerous individuals did not provide data for every available category, thus obtaining individual ROCs was problematic. Group results for the above alternative measures of sensitivity and bias are reported alongside d' and c in Tables 6.2 and 6.3 below. The different sensitivity and bias measures produce similar results, follow the same pattern across the groups, and ANOVA results for A' follow the same pattern as ANOVA results for d', Therefore, it is reasonable to argue that the present data follow the assumptions required for using d' and c in the following analyses.

Initial Analyses

As was the case in the previous chapter the current chapter assesses the underlying assumptions required for ANOVA (outliers, normality, and homogeneity of variance). Across the four age and condition groups the following dependent variables are assessed: the overall false alarm rate, d', and c; critical lure d' and c; weak lure d' and c; and unrelated d' and c.

Outliers. Outliers were identified by the same means as in the previous chapter, except Mahalanobis distances were not employed because multivariate analyses are not conducted in the present chapter. Four cases were identified as outliers for the overall false alarm rate; one younger and one older adult from the source condition, and two older adults from the standard condition. For overall d' there were two outliers; one younger adult in the standard condition, and one older adult in the source condition. Finally, there were three outliers for overall c; two older adults in the standard condition, and one older adult in the source condition.

When examining outliers for d' and c based on the different false alarm types (critical, weak, or unrelated items), no outliers were found for the critical lure d', but two were found for c; one younger and one older adult in the standard condition. For the weak lure *d′* there were three outliers; one younger adult from the standard condition, and one older adult from each of the conditions. For the weak lure *c* there were two outliers; one younger adult from the source condition, and one older adult from the standard condition. For the unrelated *d′* two outliers were found; one older and one younger adult from the standard condition. Lastly, one outlier was found for unrelated item *c*; an older adult from the source condition. As in the previous chapter two analyses were conducted for each statistical test, one included all participants' data and one excluding outliers (pairwise exclusion) from the dependent variable under analysis. Results were then compared for the two analyses. Because findings for the two sets of analyses did not differ in statistical significance, only results from the analyses containing all data are reported here.

Normality. Deviations from normality of groups' data were examined by using histograms, Q-Q plots, Kolmogorov-Smirnov normality tests, and skewness and kurtosis values and their respective *z*-scores, as in the previous chapter.

Overall false alarm rate and corresponding d′ and c. Visual inspection of histograms and Q-Q plots for the overall false alarm rate and corresponding signal detection indices indicated that many of the groups' data deviated from normality. However, the Kolmogorov-Smirnov tests showed all group's data was normally distributed (refer to Appendix M, Table M1 for the Kolmogorov-Smirnov test results). Tests revealed no groups' data was significantly skewed. For *d′* two groups had significantly leptokurtic distributions; the younger group in the standard condition and the older group in the source condition (refer to Appendix M, Table M2 for skewness and kurtosis *z*-scores).

Separate d′ and c indices. Visual examination of histograms and Q-Q plots for the critical lure, weak lure, and unrelated item *d′* and *c* indices suggested that a number of the groups' data deviated from normality. However, significant Kolmogorov-Smirnov tests were found for only the signal detection measures for unrelated items; both sensitivity and bias for younger adults in the standard condition, and bias for the younger adults in the source condition (refer to Appendix M, Table M3 for the Kolmogorov-Smirnov test results). Tests revealed the following data were significantly skewed: weak lure *d′* for older adults in the source condition, unrelated item *d′* for younger adults in the standard condition, and unrelated item *c* for younger adults in the source condition. Leptokurtic distributions were found for the following: weak lure *d′* for younger adults in the standard condition and older adults in the source condition, and unrelated item *c* for older adults in the standard condition (skewness and kurtosis *z*-scores are presented in Appendix M, Table M4).

Although a number of variables deviated from normality the magnitude of these violations was such that they were unlikely to greatly impact results (using Stevens', 2009, criteria).

Homogeneity of variance. The results of the Levene's test were significant for the critical lure *d'*, $F(3, 76) = 9.06$, $p < .001$, and unrelated item *c*, $F(3, 76) = 2.97$, $p = .04$. Fortunately, as discussed previously, the equal group sizes in the present research means that *F* can be assumed to be robust (Cardinal & Aitken, 2006; Stevens, 2009).

Adjusted Alpha Level

Similar to Chapter 5, the present chapter employed the Holm (1979) procedure to calculate family-wise error. In an attempt to account for inflated Type I error, whilst not being overly conservative, six families of significance tests were defined as follows. The two-way between groups ANOVA conducted on the overall false alarm rate was one family, as was the ANOVA conducted on *d'*, and *c*, both based on the overall false alarm rate – these three ANOVAs were considered separate families because they were the main analyses of the present chapter (each consisted of three significance tests). The fourth family consisted of the two ANOVAs (six significance tests) and pairwise comparisons (six significance tests) employed to examine group differences for critical lure *d'* and *c*. Another family consisted of the two ANOVAs conducted on the weak lure *d'* and *c* (six significance tests). The final family was the two ANOVAs that examined the unrelated item *d'* and *c* (six significance tests). As in the previous chapter, the current chapter will present unadjusted and adjusted significance values (*p* and *p'*, respectively), and present adjusted alpha levels (α') only when $p < .001$.

Sensitivity and Bias: Overall False Alarm Fate and Hit Rate

Below are the results from three separate two-way between-groups ANOVAs. The first ANOVA examines group differences for the overall false alarm rate. The second and third ANOVAs examine group differences on the sensitivity and criterion indices (respectively). The group means for these three measures can be found in Table 6.2.

Table 6.2

Group Means for Measures Based on the Overall False Alarm Rate

	Young		Old	
Variable	Standard task	Source task	Standard task	Source task
Hit rate	.69 (.09)	.71 (.10)	.69 (.17)	.59 (.20)
False alarm rate	.33 (.17)	.23 (.11)	.40 (.19)	.28 (.15)
Sensitivity (*d'*)	1.01 (.37)	1.37 (.42)	.83 (.35)	.93 (.32)
Sensitivity (*A'*)	.77 (.08)	.82 (.06)	.73 (.08)	.75 (.06)
Sensitivity (A_g)	.72	.78	.67	.67
Criterion (*c*)	-.02 (.36)	.12 (.28)	-.14 (.51)	.19 (.50)
Criterion (c_2)	-.02	.09	-.10	.15

Overall false alarm rate. The two-way ANOVA conducted to examine group differences on the overall false alarm rate demonstrated that there were no age group differences (F[1, 76] = 2.93, p = .09, p' = .18, η^2 = .03, refer to Figure 6.1 below). Yet, irrespective of age, those in the standard recognition condition had a higher false alarm rate than participants in the source condition (M = .36, SD = .18 and M = .25, SD = .13, respectively. Refer to Figure 6.2 below), F(1, 76) = 10.42, p = .002, p' = .006, η^2 = .12. The interaction was not significant (F[1, 76] = 0.08, p and p' = .78, η^2 < .01), indicating that the change in test conditions impacted older and younger adults similarly.[10]

Overall sensitivity. The two-way ANOVA conducted to examine *d'* showed that, regardless of the test condition, older adults (M = 0.88, SD = .33) were less sensitive than younger adults (M = 1.19, SD = .43), F(1, 76) = 14.24, p < .001, α' = .02, η^2 = .14. Those in the source-monitoring condition (M = 1.15, SD = .43) had greater sensitivity than those in the standard condition (M = 0.92, SD = .37), F(1, 76) = 7.65, p = .007, p' = .01, η^2 = .08. These results can be seen in Figures 6.3 and 6.4 below. No interaction effect was found (F[1, 76] =

[10] For the test conditions, group differences were found for both overall *d'* and *c*. Both bias and sensitivity can impact the false alarm rate. Therefore, the test condition difference in the overall false alarm rate must be interpreted with caution, as it is unclear which (sensitivity or bias) is producing the difference.

2.55, p and p' = .11, η^2 = .03), indicating that the change in test conditions impacted older and younger adults' sensitivity equally.

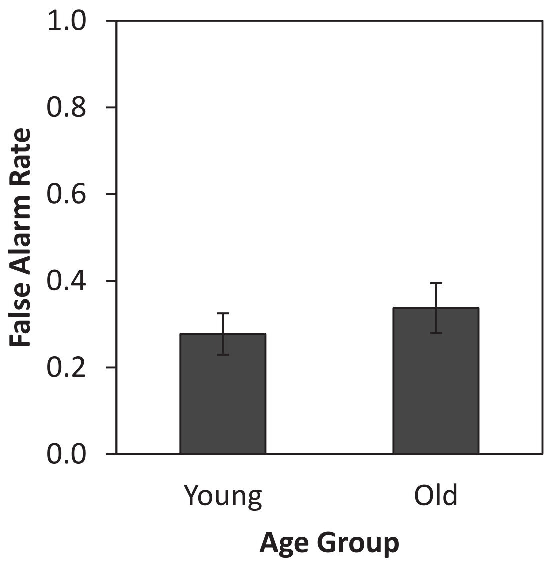


Figure 6.1. Mean overall false alarm rate across the age groups. Vertical lines show 95% CIs.

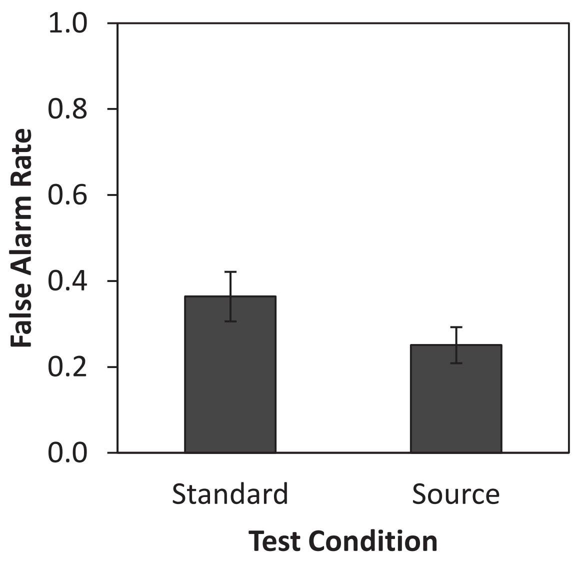


Figure 6.2. Mean overall false alarm rate across the test conditions. Vertical lines show 95% CIs.

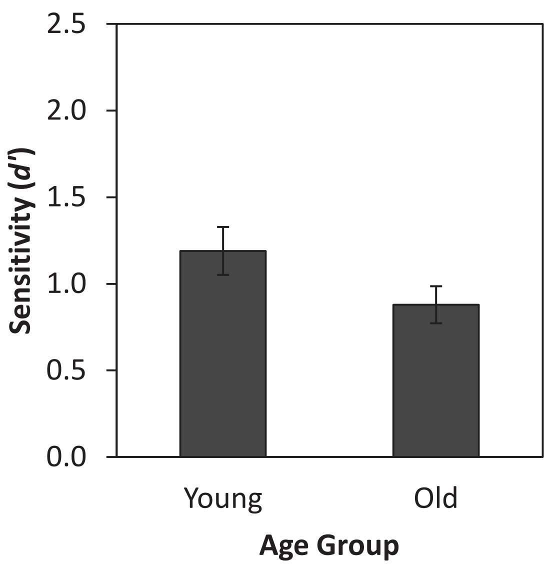


Figure 6.3. Mean sensitivity (based on the overall false alarm rate) across the two age groups. Vertical lines show 95% CIs.

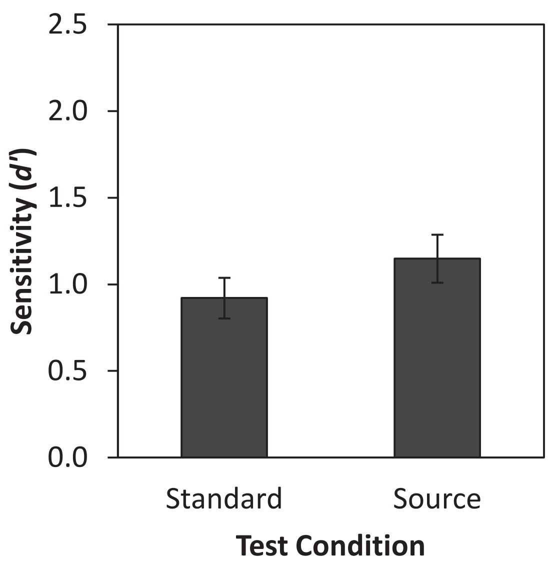


Figure 6.4. Mean sensitivity (based on the overall false alarm rate) for the two test conditions. Vertical lines show 95% CIs.

Overall bias. As can be seen in Table 6.2 both age groups in the standard condition were biased to respond old, while both age groups in the source condition were biased to respond new. The results from the two-way ANOVA for *c* found no differences between age groups, *F*[1, 76] = 0.07, *p* and *p′* = .79, $\eta^2 < .01$, both of which were biased to respond new, as can be seen in Figure 6.5 below. By contrast, those in the source condition (*M* = .16, *SD* = .40) used a stricter criterion (and were biased to respond old) than those in the standard condition (*M* = -.08, *SD* = .44), *F*(1, 76) = 6.03, *p* = .02, *p′* = .05, $\eta^2 = .07$ (who were biased to respond new – see Figure 6.6 below). No interaction effect was found (*F*[1, 76] = 1.04, *p* = .31 *p′* = .62, $\eta^2 = .01$), indicating older and younger adults used a similar criterion, and the change in test condition influenced both age groups similarly.

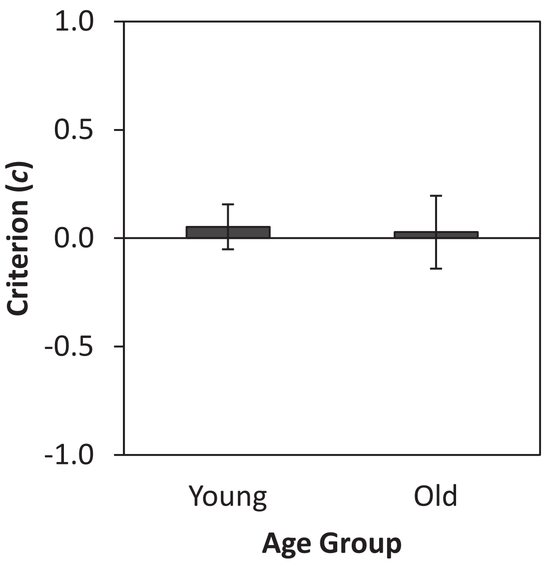


Figure 6.5. Mean criterion (based on the overall false alarm rate) across the two age groups. Vertical lines represent 95% CIs.

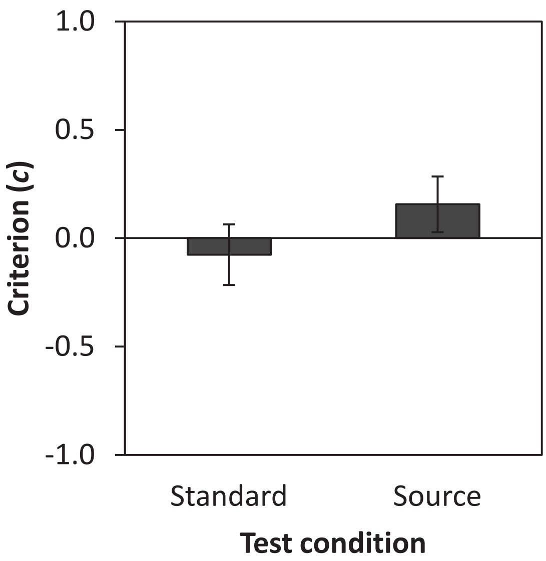


Figure 6.6. Mean criterion (based on the overall false alarm rate) for the two test conditions. Vertical lines represent 95% CIs.

Sensitivity and Bias: Separate Types of False Alarm Rates and Hit Rate

To further examine group differences separate sensitivity and bias indices were calculated using each of the three separate false alarm types and the hit rate. Separate two-way ANOVAs were then conducted to find significant group differences. Table 6.3 provides the means and standard deviations for the different false alarm rates and variables employed in the following analyses.

Table 6.3

Group Means for Measures Based on the Separate Item Type False Alarm Rates

	Young		Old	
Variable	Standard task	Source task	Standard task	Source task
Critical lure				
False Alarm Rate	.73 (.15)	.49 (.27)	.79 (.15)	.72 (.18)
Sensitivity (*d'*)	-.19 (.40)	.62 (.95)	-.38 (.39)	-.43 (.45)
Sensitivity (*A'*)	.41 (.15)	.63 (.24)	.33 (.11)	.34 (.14)
Sensitivity (A_g)	.47	.64	.43	.42
Criterion (*c*)	-.60 (.36)	-.24 (.4)	-.73 (.49)	-.48 (.53)
Criterion (c_2)	-.62	-.30	-.74	-.44
Weak lure				
False Alarm Rate	.33 (.22)	.26 (.14)	.43 (.23)	.25 (.18)
Sensitivity (*d'*)	1.03 (.56)	1.25 (40)	.71 (.53)	1.09 (.36)
Sensitivity (*A'*)	.76 (.11)	.81 (.06)	.69 (.15)	.79 (.05)
Sensitivity (A_g)	.74	.77	.66	.70
Criterion (*c*)	.01 (.45)	.07 (.33)	-.18 (.60)	.28 (.59)
Criterion (c_2)	-.02	.05	-.14	.19
Unrelated item				
False Alarm Rate	.14 (.13)	.08 (.06)	.16 (.18)	.12 (.12)
Sensitivity (*d'*)	1.76 (.52)	2.03 (.43)	1.71 (.49)	1.60 (.57)
Sensitivity (*A'*)	.87 (.07)	.90 (.03)	.87 (.07)	.85 (.07)
Sensitivity (A_g)	.83	.87	.79	.75
Criterion (*c*)	.38 (.39)	.47 (.26)	.32 (.52)	.54 (.49)
Criterion (c_2)	.27	.36	.21	.37

Note. False alarm rates in this table are the transformed false alarm rates.

Critical lure sensitivity and bias. The two-way ANOVA conducted on critical lure *d’* showed a significant difference between the two age groups, $F(1, 76) = 21.58$, $p < .001$, $\alpha' =$ .004, $\eta^2 = .19$, and the two test conditions, $F(1, 76) = 8.28$, $p' = .03$, $p = .005$, $\eta^2 = .07$. There was also a significant interaction, $F(1, 76) = 10.38$, $p' = .02$, $p = .002$, $\eta^2 = .09$. Pairwise comparisons (as recommended by Pallant, 2007; Tabachnick & Fidell, 2013) demonstrated that sensitivity of the younger adults in the source-monitoring condition (*M* = 0.62, *CI* [0.36, 0.88]) was significantly better than the three other groups ($p < .001$ and $\alpha' = .005$, for the three comparisons), who did not differ in sensitivity from each other; that is, younger adults in the standard condition (*M* = -0.19., *CI* [-0.46, -0.07]), older adults in the standard condition (*M* = -0.38, *CI* [-0.64, -0.12]), and older adults in the source-monitoring condition (*M* = -0.43, *CI* [-0.69, -0.61]) had equal sensitivity. Figure 6.7 below displays these findings.

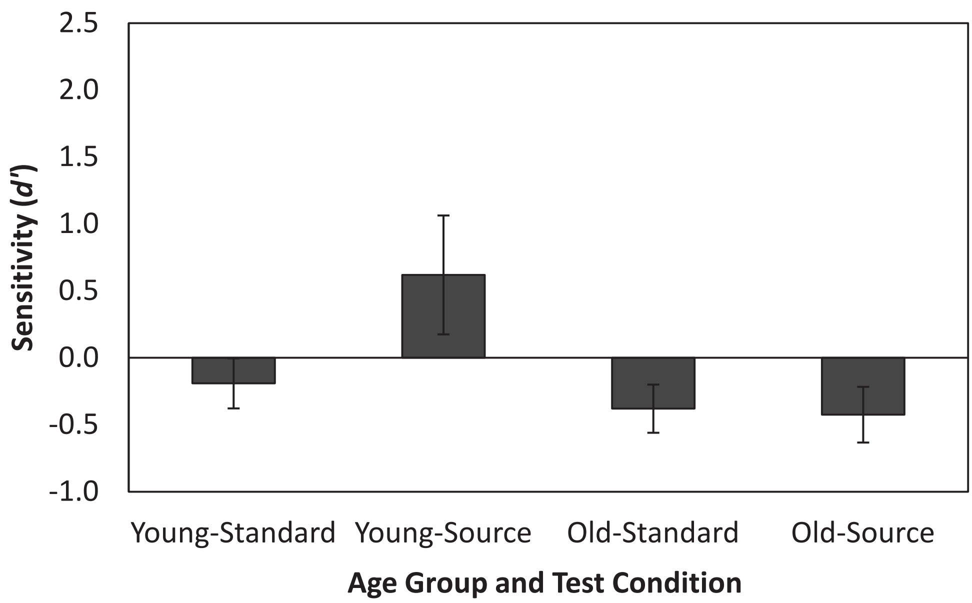


Figure 6.7. Mean sensitivity (based on false alarms to critical lures) across the four age and condition groups. Vertical lines represent 95% CIs.

For critical lure *c*, there were no age group differences, $F(1, 76) = 3.16$, $p' = .08$, $p =$ .40, $\eta^2 = .04$ (see Figure 6.8 below). Yet, those in the source condition (*M* = -0.36, *SD* = .48) had a significantly more conservative criterion than those in the standard condition (*M* = -0.66, *SD* = .42), $F(1, 76) = 9.14$, $p' = .02$, $p = .003$, $\eta^2 = .10$ (see Figure 6.9 below). As can be seen in Table 6.3 above, at test all four groups favoured responding that they thought critical lures were studied (i.e., old). Lastly, the interaction was not significant, $F(1, 76) = 0.31$, $p' =$

1.0, p = .58, η^2 = .06, suggesting that the two test conditions impacted older and younger adults similarly.

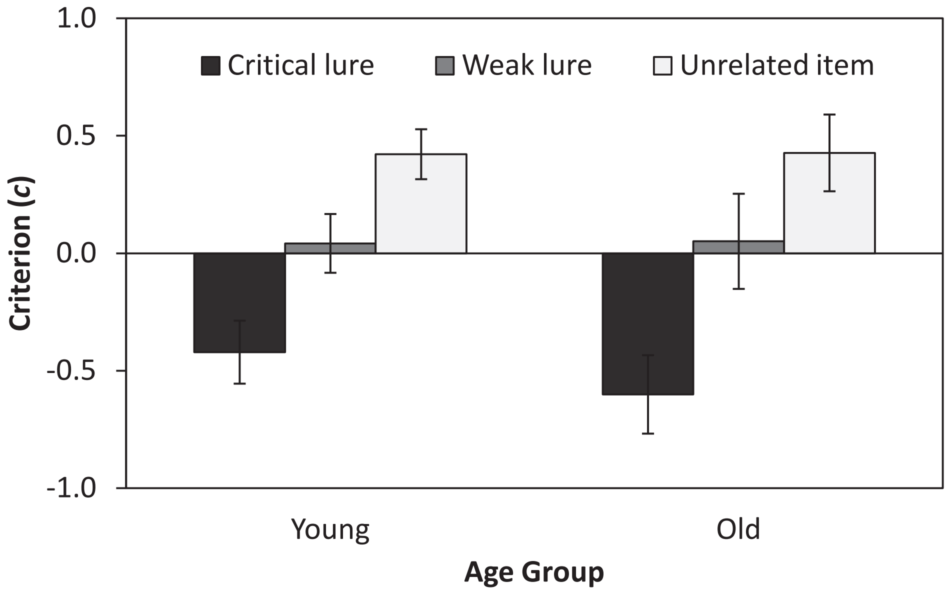


Figure 6.8. Mean critical lure, weak lure, and unrelated item criterion indices across the two age groups. Vertical lines represent 95% CIs.

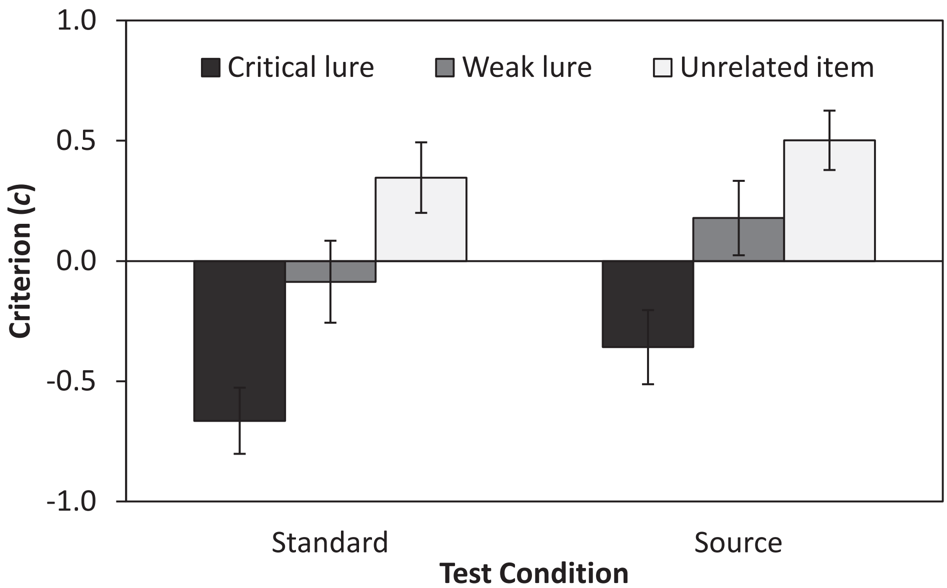


Figure 6.9. Mean critical lure, weak lure, and unrelated item criterion indices across the two test conditions. Vertical lines represent 95% CIs.

Weak lure sensitivity and bias. For weak lure *d'*, the two-way ANOVA results found no significant differences between age groups, $F(1, 76) = 5.21$, $p' = .11$, $p = .03$, $\eta^2 = .06$ (see Figure 6.10 below). Those in the source-monitoring condition ($M = 1.17$, $SD = .38$) had greater sensitivity than those in the standard condition ($M = .87$, $SD = .56$), $F(1, 76) = 8.14$, $p' = .04$, $p = .006$, $\eta^2 = .09$ (see Figure 6.11 below). The interaction effect was not significant, $F(1, 76) = 0.70$, $p' = .83$, $p = .42$, $\eta^2 = .01$.

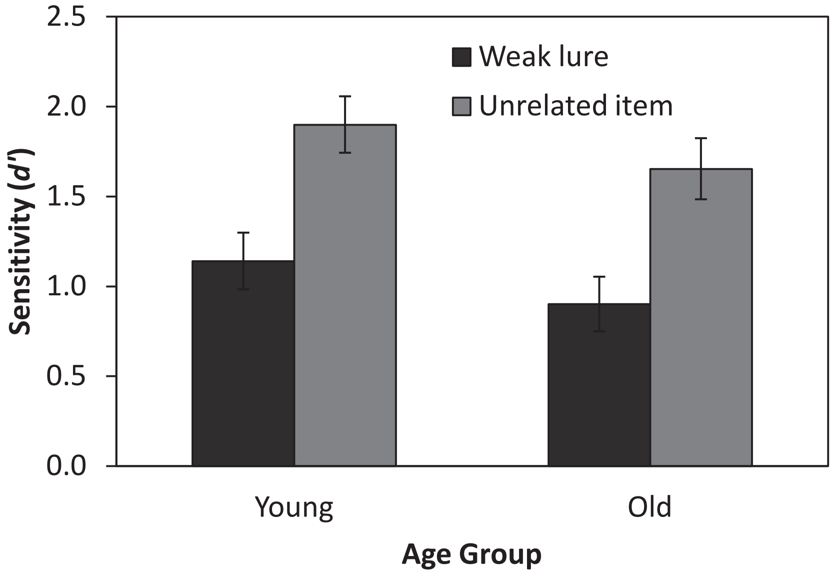


Figure 6.10. Mean weak lure and unrelated item sensitivity across the two age groups. Vertical lines represent 95% CIs.

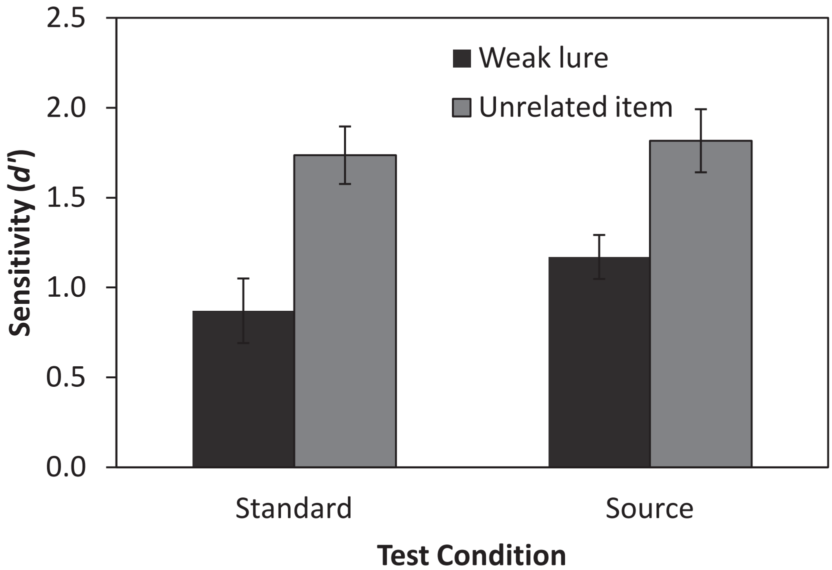


Figure 6.11. Mean weak lure and unrelated item sensitivity across the two test conditions. Vertical lines represent 95% CIs.

For weak lure *c* the descriptive statistics suggest that the older adults in the standard condition were biased to respond old, whilst older adults in the source condition were biased to respond new, and both younger groups' bias was near zero, indicating little or no bias. However, the two-way ANOVA for weak lure *c* found no differences between age groups, test conditions, or an interaction ($F[1, 76] = 0.01$, $p' = .93$, $p = .94$, $\eta^2 = < .01$; $F[1, 76] = 0.02$, $p' = .11$, $p = .02$, $\eta^2 = .06$; and $F[1, 76] = 0.08$, $p' = .24$, $p = .08$, $\eta^2 = .04$, respectively). See Figures 6.8 and 6.9.

Unrelated item sensitivity and bias. No significant effects were found for unrelated item *d'* or *c* (Figures 6.8 to 6.11). As can be seen in Table 6.3, all groups were bias to respond new. ANOVA results for *d'* were as follows: age effect, $F(1, 76) = 4.70$, $p' = .20$, $p = .03$, $\eta^2 = .06$; test condition effect, $F(1, 76) = 0.50$, $p' = 1.0$, $p = .48$, $\eta^2 < .01$; and interaction, $F(1, 76) = 2.81$, $p' = .49$, $p = .10$, $\eta^2 = .03$. ANOVA results for *c* were as follows: age effect, $F(1, 76) = 0.01$, $p' = 1.0$, $p = .95$, $\eta^2 < .01$; test condition effect, $F(1, 76) = 2.62$, $p' = .49$, $p = .11$, $\eta^2 = .04$; and interaction, $F(1, 76) = 0.42$, $p' = 1.0$, $p = .52$, $\eta^2 < .01$.

Summary

In this chapter, as in the previous one, data from a DRM false memory task from two age groups (aged 16-30 and 75-80) and two test conditions (standard old/new recognition task and source-monitoring task) were examined. The previous chapter utilised multivariate and univariate analyses of variance to examine group differences in recognition rates (false alarm rates), the present chapter examined signal detection measures of sensitivity and response bias. Firstly, false alarm scores across the three false items were combined and examined. The ANOVA revealed no age-related differences for analyses that used the overall false alarm rate, whilst those in the standard test condition had a higher overall false alarm rate than those in the source-monitoring test condition. Older adults were less sensitive than younger adults, and those in the standard condition were less sensitive than those in the source condition. There were no age-related differences for bias (calculated using the overall false alarm rate and the hit rate) with both age groups being biased to respond new/unheard. Those in the standard test condition were biased to respond old/heard, while those in the source condition were biased to respond new/unheard. No significant interactions were found for the overall false alarm rate, sensitivity, or bias, indicating the different test conditions impacted age group performance equally on these measures.

Separate sensitivity and bias indices were also calculated using the different item types' false alarm rates (giving three separate sensitivity and bias indices for critical lures, weak lures, and unrelated items). For sensitivity and criterion indices based on critical lure false alarms, analyses revealed that the younger source-monitoring group had better sensitivity than the other three groups (who all had similar sensitivity). No age-related differences were found for bias indices, whilst the standard condition was found to be more liberal than the source-monitoring condition (all groups were biased towards old/heard responses). The two test conditions impacted the two age groups bias indices similarly.

Analyses of sensitivity and criterion measures based on weak lure false alarms found no age group differences for either sensitivity or bias. When comparing test conditions the standard condition was found to be less sensitive than the source condition. The examination of the mean bias indices across groups revealed that the older adults in the standard test condition were biased to response old/heard, both of the younger groups' bias indices were near zero (no bias), whilst the older adults in the source-monitoring condition were biased to respond new/unheard. The lack of interactions indicated sensitivity and bias was impacted similarly by the different test conditions. Finally no group differences were found for sensitivity and bias indices that were based on unrelated item false alarms, and all groups were biased to respond new/unheard.

Chapter Seven

Discussion

Most people would prefer to believe that their memories are true representations of the past. However, false memories are a common phenomenon, and research has shown such memories can be easily created with certain tasks. Additionally, years of research have demonstrated older adults are sometimes, but not always, more prone to false memories than younger adults. The present research attempted to provide deeper understanding of the effects of ageing on the creation of false memories, using the DRM (Deese/Roediger-McDermott) paradigm and a source-monitoring manipulation found to improve recognition performance (Multhaup & Conner, 2002), with the aim of establishing if differences in false recognition are best explained by (a) fuzzy-trace theory, (b) source-monitoring processes, or (c) criterion and/or sensitivity differences in signal detection ability. In the current research the recognition test included not only critical lures, but also weak lure words, which have been shown to be a more sensitive measure of false recognition. Lastly, signal detection theory was utilised to provide further information about participants' sensitivity and bias when making recognition decisions in a false memory task.

This chapter discusses the results and implications of the present study that have been outlined above. Firstly, findings related to source memory are discussed, followed by general false memory results. Secondly, the results are discussed and explained using the activation-monitoring theory, followed by the fuzzy-trace theory. Next, criterion and sensitivity differences are discussed regarding how they explain the findings. The discussion then turns to the implications and applications the present research's finding have to real life and everyday false memories. Lastly, the limitations of the current research and suggestions for future research are discussed.

Source Memory

In requiring participants to use a source-monitoring strategy (in the source-monitoring test condition) the expectation was that memory performance would be more accurate compared to a standard yes/no recognition task (Hicks & Marsh, 1999; Multhaup, 1995; Multhaup & Conner, 2002). The present study's results support this prediction, with the source-monitoring group having lower false recognition rates than those in the standard test condition for critical lures, weak lures, and the overall (critical lures, weak lures, plus unrelated items) measure. Furthermore, those in the source-monitoring condition had better sensitivity for both the overall false item measure and weak lures. These findings may mean that source-monitoring instructions encouraged participants to use item-specific processing, because source decisions require judgements to be made based on the differences between memory characteristics of sources and/or similarities between memory characteristics and activated schemas (M. K. Johnson et al., 1993).

As discussed in Chapter 2, both the activation-monitoring theory and the fuzzy-trace theory embrace the idea that individuals hold a set of rules about the information they need to retrieve in order to decide an item was present at study (Brainerd & Reyna, 2002a, 2005). Requiring participants to monitor the source of items is expected to encourage the use of more conservative rules compared to situations in which source-monitoring is not required (Multhaup, 1995; Multhaup & Conner, 2002). Similar to research by Multhaup (1995) and Multhaup and Conner (2002), the inclusion of a source-monitoring task caused participants to use a more conservative acceptance criterion for both overall false alarms and critical lures alone. Although this result was expected, it was quite different from Hicks and Marsh's (1999; 2001; described earlier) findings, where the source task led participants to use a more liberal criterion and increased false recognition when compared to a standard condition. The present study's results further support the use of source tasks that include the correct source(s) in the response options to decrease the incidence of false memories.

Source confusion is another idea common to both the activation-monitoring theory and the fuzzy-trace theory. Source confusions are believed to occur because source-identifying information becomes disconnected from studied items and incorrectly associated with non-presented items (Reyna & Lloyd, 1997). The high false recognition of critical lures, even after requiring participants to source monitor, suggests that critical lures were strongly activated in participants' semantic memory (activation-monitoring theory) or associated gist representations (fuzzy-trace theory). The high false recognition rate may indicate that source information became disconnected from the presented items and incorrectly associated with the critical lures (as outlined by the binding hypothesis discussed earlier). In addition, older adults incorrectly categorised critical lures as old more often that younger adults, suggesting older adults had greater difficulty storing and retrieving source information associated with recently experienced items. This is in line with the findings of other researchers who state older adults have problems binding source memory characteristics correctly (Dodson et al., 2007; Ferguson et al., 1992; Henkel et al., 1998).

False Memories

The previous section focused on performance differences between the two test conditions and ideas about source that are common to both the fuzzy-trace and activation-monitoring theory. The current section moves away from the source findings and discusses the false memory findings and how they compare to previous research. The present research provides further evidence of the robustness of the DRM false memory phenomenon. One striking result was that three of the four age and condition groups categorised critical lures as old more frequently than they did old items. Only the younger adults in the source-monitoring condition correctly categorised a higher proportion of presented items as old than critical lures as old. This is striking because, usually, false recognition of lures occurs at a lower (Gallo, 2006, discussed earlier) or equal (Mather, Henkel, & Johnson, 1997; Payne et al., 1996; Roediger & McDermott, 1995; Tun et al., 1998) rate compared to correct recognition of studied items. Another striking finding was how certain both age groups were that falsely recognised critical lures *had* been studied at test; certainty ratings averaged somewhere between *very certain heard* and *mostly certain heard*.

Unlike the research reviewed by Gallo (2006; in which data from 15 standard DRM recognition experiments was combined) that demonstrated age-related differences in correct recognition, but not false recognition (with less than half of the experiments finding significant differences between age groups), the standard condition of the current research found no age group differences for correct recognition (hits), and an age group difference for false recognition of critical lures, but not weak lures. The present experiment cannot be easily compared to the standard DRM recognition tasks reviewed by Gallo, as it included weak lures at test. Weak lures were expected to provide a more sensitive way of measuring age-related differences in false recognition. As discussed earlier, their inclusion should make thematic associations inefficient as the sole basis for making a recognition decision (which relates both to activation and gist mechanisms). As a result age-related differences should be larger for weak lures compared to critical lures (Tun et al., 1998). Similar to Tun et al.'s findings the present research found older adults had higher false recognition of critical lures than younger adults but there was no difference between the two groups for correct recognition. Unlike Tun et al.'s finding the present research found no age-related differences in false recognition of weak lures.

The difference between Tun et al.'s (1998) findings and the present study was likely due to the fact that Tun et al. tested participants with a recall test followed by a recognition test immediately after presenting each of the ten DRM lists. By contrast, in the present research the recognition test was given after all eight DRM lists had been presented. A review by Gallo (2006) found larger age group differences result when recognition tests are confounded by prior recall testing compared to when they are not. The use of prior recall testing may have meant that, compared to the present research, there was an additional opportunity for weak lures to become semantically activated or part of one's gist representation, and that there was an additional source to be confused with the study phase. These three factors would increase the likelihood weak lures would be falsely recognised by participants of any age, but they would have a more adverse impact on older than younger adults because older adults have poorer source memory (e.g., Henkel et al., 1998; McDaniel et al., 2008; Rosa & Gutchess, 2011; Spencer & Raz, 1995), rely more on gist than verbatim traces, and have difficulty using verbatim traces to offset the accumulation of gist traces (Budson et al., 2000; Kensinger & Schacter, 1999). It is possible that, when the DRM task does not include recall testing prior to recognition testing, weak lures do not provide a more useful measure of age-related differences in false recognition than critical lures.

The Activation-Monitoring Theory

The previous sections have focused on general source memory and false memory findings for the current research. This section focuses on how the research findings can be explained by the activation-monitoring theory (the next section will do the same for fuzzy-trace theory). Activation-monitoring theory predicts that older adults will perform worse than younger adults in false recognition tasks (higher false recognition *and* lower correct recognition) because source memory and recognition memory are impaired (Ferguson et al., 1992; Hashtroudi et al., 1990; McIntyre & Craik, 1987; Schacter et al., 1991), while semantic activation remains stable with ageing (Balota et al., 1999; Dehon & Brédart, 2004; Tun et al., 1998). The present research found results that both do and do not fit this expectation.

Compared to younger adults, older adults falsely recognised more critical lures, and were more certain critical lures had been presented. These recognition findings suggests that *if* participants employed a monitoring strategy, older adults' monitoring ability was impaired compared to younger adults. Further support for older adults' monitoring ability being impaired is that the different recognition rates showed that only source-monitoring by younger adults in the source-monitoring condition was sufficiently effective to lead to lower false recognition of critical lures compared to correct recognition of presented items. A ceiling effect may have contributed to the three other groups' higher recognition of critical lures compared to presented items. This is discussed further in the limitations section below.

In contrast to the critical lure results, no age group differences were found for false recognition of weak lures, which are meant to be a more sensitive measure of false recognition (Tun et al., 1998). This finding suggests that, although there was an age-related impairment in the source-monitoring of critical lures, the monitoring strategy was as effective for older adults as it was for younger adults for weak lures. The signal detection analyses also support this interpretation, as the weak lure sensitivity index did not differ between age groups, indicating that the monitoring strategy effectively discriminated weak lures from presented items for both age groups. The difference in the effectiveness of the monitoring strategy for critical lures and weak lures was likely due to the fact that the latter were not as strongly activated as the former. It is clear from the high recognition rates of critical lures that these items were strongly activated, and all four groups' false recognition rates follow the pattern that would be expected due to the different strength in semantic activation of the item types (critical lures > weak lures > unrelated items). Interestingly, older adults were more certain than younger adults that falsely recognised weak lures had been presented. This finding suggests that although the strategy was effective for both age groups to identify weak lures as new, it was still somewhat impaired for older adults.

As discussed earlier, Gallo (2006) states that limited item-specific differences and semantic variation between presented and non-presented items make it difficult to formulate accurate judgements about whether an item was presented during study. Although this may be the case for critical lures, it appears the opposite was the case for weak lures; that is, the item-specific differences and amount of semantic variation between items allowed accurate judgements to be made. This also appeared to be the case for unrelated words; however, this finding, along with the finding of no group differences in false recognition of unrelated words, is as expected, because unrelated items have no semantic relationship with presented words. Group differences in false recognition of unrelated items are rarely found, and results for unrelated items tend not to be discussed in publications. Furthermore, the finding that older and younger adults correctly identified weak lures and unrelated words equally well shows that, despite research consistently finding that older adults have poorer source-memory than younger adults, in some circumstances (i.e., when item-specific features and semantic variation greatly differ for presented items and non-presented items) older adults do perform as well as younger adults.

No age-related differences were found for recognition of presented items. This result is unexpected considering older adults generally perform more poorly compared to younger adults on recognition tasks due to their general memory performance impairments related to cognitive ageing. This result suggests that the underlying mechanisms or strategies involved in normal recognition memory are different from that of false memory tasks. Finding no age group differences for recognition of presented items while finding age group differences in false recognition, implies that in the DRM task the mechanisms at work include not only an effortful controlled process (e.g., source-monitoring), but also a strategy based on automatic influences (e.g., semantic activation and familiarity). This is because effortful controlled processes would result in age-related differences in recognition, but automatic processes would results in similar recognition across age groups. Together, the above age-related findings support the claim that activation *and* monitoring mechanisms underpin false memory task performance. Further support for this interpretation is the finding that recognition rates across the item types (critical lures, weak lures, presented items, and unrelated items) were different within each age and condition group. In particular, false recognition of weak lures was lower than recognition of presented items for all age and condition groups. If a monitoring strategy had not been used, similar recognition rates for presented items and weak lures would have resulted, because the strength of the semantic associations (and thus activation) amongst presented items and weak lures is similar.

It is thought that when a DRM task does elicit a spontaneous monitoring strategy (which is sometimes the case; Gallo, 2006) older adults are less likely than younger adults to use such a strategy because monitoring is an effortful and controlled process, and older adults' ability to use such processes is often impaired (Jacoby, 1999; Jacoby et al., 1996; Jennings & Jacoby, 1993, 1997; Skinner & Fernandes, 2009). If older adults do not use a source-monitoring strategy in the standard condition, when they are *forced* to source-monitor, their performance should improve considerably more than that of younger adults. This prediction was not supported by results from the current study. The lack of any significant interactions in the two-way ANOVAs for recognition rates shows that the source-monitoring task improved performance of both age groups similarly. This result provides further evidence that both age groups used the same strategy (possibly source-monitoring) in both the standard and source conditions, but younger adults' strategy was somewhat more effective, as they were better than older adults at identifying critical lures as new items, and at categorising critical lures using the source task response options. Therefore, forced source-monitoring enhanced an already used monitoring strategy. This outcome differs from expectations of other authors who believe older adults are more inclined to rely on familiarity or activation mechanisms to make recognition decisions, whereas younger adults employ a monitoring mechanism to offset familiarity via activation (Skinner & Fernandes, 2009).

The Fuzzy-Trace Theory

According to the fuzzy-trace theory gist and verbatim traces are both responsible for true memories, but they work in opposition for false memories, as verbatim traces suppress false memories by counteracting the influence of gist traces. Research shows that older adults exhibit higher false recognition than younger adults for items semantically related to presented items (i.e., weak lures and critical lures, and thus the overall false recognition measure), while correct recognition will be similar across age groups. According to the fuzzy-trace theory this is because older adults are believed to rely more on gist traces than younger adults do (Budson et al., 2000; Kensinger & Schacter, 1999); this cognitive state of affairs is likely to impact false recognition but not true recognition, as correct recognition can be based on gist and/or verbatim traces. Just as was the case for the activation-monitoring model, the present study's results provide evidence for and against this prediction.

Consistent with the fuzzy-trace theory, recognition of presented items was equal across age groups. If older adults have difficulty with verbatim traces, this result is consistent with the idea that older adults are able to rely more than younger people on gist traces to compensate for their impairment with verbatim traces. As expected, age-related differences were found for false recognition of critical lures, suggesting that older adults were more reliant on gist traces than younger adults. Older adults were more certain than younger adults that critical lures, weak lures, and presented items were heard, which again suggests older adults were more reliant on gist memories than younger adults. This is the case because reliance on gist memories likely corresponds to holding a higher certainty that false items were presented, as new items semantically related to old items are sufficiently similar to the gist traces of the old learnt items that the new test items are considered old. By contrast, reliance on verbatim as well as gist memories leads to less certainty that false items were presented because verbatim details cannot be recollected for false memories. In the current study age-related differences were not found for weak lures. These results are quite unexpected from a fuzzy-trace point of view and difficult to explain. They suggest that older adults were *not* more reliant on gist traces than younger adults, because if older adults had been the study would have found older adults had higher false recognition compared to younger adults for critical lures and weak lures, not just critical lures.

One possible explanation for the age-related findings for weak lures (no age difference) and critical lures (age difference) is that the reliance on gist and verbatim traces differs for the different item types. As discussed previously, presented items cue verbatim more frequently than gist traces, thus correct recognition is thought to rely more heavily on the retrieval of verbatim details than gist details. By contrast, false items cue gist traces; thus, false recognition is based on the retrieval of gist information (Brainerd & Reyna, 2002a). The weaker the relationship between false items and the overall theme of the DRM list, the less likely they will cue gist traces. Critical lures are excellent cues for gist traces, as evidenced by participants' high misrecognition of them, whereas weak lures are not as effective as critical lures at cueing gist traces. Because false recognition is due to the retrieval of gist information, and correctly rejecting a lure requires verbatim processes, older adults should still misrecognise more weak lures than younger adults due to their impairment of verbatim memory processes. Verbatim processes that allow one to correctly reject a lure include retrieval of verbatim traces that counteract the perceived familiarity of false items, and using a strict set of rules in the decision process that mean false items will be rejected because clear verbatim details of the experience are not available. No age-related differences were found for the bias measures, suggesting that if participants employed rules regarding verbatim recollection they were the same across the two age groups.

To put these results more simply, if the present study's results had followed the general pattern outlined by the fuzzy-trace theory, older adults would have had higher false recognition for critical lures, weak lures, and overall false recognition. Such results were not found. Furthermore, if the false memory task had resulted in the similar use of gist processes by both age groups, age-related differences in false recognition would not have been found. Yet, there was an age group difference for false recognition of critical lures. By contrast, if participants relied on both gist and verbatim traces the study would have found older adults' recognition performance was worse than younger adults' for critical lures, weak lures, and overall false recognition. Age differences were not found for weak lures. These findings are difficult to reconcile with the claims of fuzzy-trace theory.

From a fuzzy-trace perspective the source-monitoring manipulation may have encouraged participants of both age groups to be less reliant on gist memories and/or more reliant on verbatim memories compared to the standard test condition. Unlike the activation-monitoring theory, which predicts requiring source-monitoring will decrease the incidence of false recognition considerably more in older adults than in younger adults, the fuzzy-trace theory predicts that the incidence of false recognition in older adults will decrease, but not considerably more than for younger adults. Again, this is due to older adults' impaired verbatim memory, because successful use of source-monitoring to offset the impact of gist memories requires good verbatim memory (Budson et al., 2000; Kensinger & Schacter, 1999). This is because, according to the fuzzy-trace theory, intact verbatim traces are required to avoid misrecognising lures, and the source-monitoring instructions cannot improve verbatim processing. Including a source-monitoring task with the DRM task improved older and younger adults' memory performance equally for all of the recognition measures (critical lures, weak lures, overall false recognition, presented items, and unrelated items). This suggests that when required to source-monitor, older adults were able to use verbatim traces (focusing more on item-specific details) to offset gist memories as well as younger adults to improve their performance. The lack of any interactions for certainty ratings for the four item types also provides evidence for this interpretation.

Signal Detection

The two sections above have shown that both the activation-monitoring theory and the fuzzy-trace theory can adequately explain many, but not all, of the findings from the current research. The previous sections have also included many of the signal detection results to support the interpretations discussed. The current section discusses the sensitivity and response bias results, focusing on how the signal detection analyses added to the information provided by the basic recognition accuracy analyses. To ease understanding Table 7.1 below provides a summary of whether or not age and/or condition group differences were found for the different true and false memory measures across the signal detection and recognition accuracy analyses.

Table 7.1

Summary of Significant Differences from the Recognition Accuracy and Signal Detection Analyses for the Age Groups and Test Conditions

	Recognition Accuracy	Sensitivity	Response Bias
Age difference			
Overall false alarms	No	Yes	No
Critical lures	Yes	Yes[a]	No
Weak lures	No	No	No
Unrelated items	No	No	No
Test condition difference			
Overall false alarms	Yes	Yes	Yes
Critical lures	Yes	Yes[a]	Yes
Weak lures	Yes	Yes	No
Unrelated items	No	No	No

[a]These results refer to the interaction effect which showed younger adults' in the source condition sensitivity differed from the other three groups.

One measure that the standard recognition analyses do not provide is response bias. The results of the present study found no age related differences for response bias. However, those in the source-monitoring condition were found to respond more conservatively (both for analyses that used all false alarms and those that used only the critical lures) than those in the standard condition. The source-monitoring task appears to make people respond more cautiously, even though it does not explicitly request people to be more cautious. Multhaup (1995) argues that it may be that requiring a source decision within a false recognition task fosters a reasonably stringent acceptance criterion, even without an explicit request to respond cautiously. Researchers do not tend to use such explicit instructions in false memory tasks, unless they are using conditions which provide warnings about the creation of false memories and/or how to avoid them (e.g., Gallo et al., 1997; McCabe & Smith, 2002; Neuschatz et al., 2003; Watson et al., 2004). Interestingly, participants in the source-monitoring condition were not simply more conservative in general. If this were the case then the hit rate would also be lower, but hit rates were similar across conditions. This result suggests that older and younger adults in the source-monitoring condition were better able to adopt a criterion that minimised false alarms and maximised hits compared to those in the standard recognition condition.

In line with Multhaup's (1995) view that requiring a source decision promotes the use of a reasonably stringent acceptance criterion is the idea that source decisions require participants to inspect their memory of an item more carefully (Hicks & Marsh, 1999; Multhaup & Conner, 2002). In the present research those in the source condition were found to have better sensitivity (both for analyses that used all false alarms and those that used only weak lures). These results may indicate those in the source-monitoring condition inspected their memories more carefully than those in the standard condition. False memory tasks that include a source judgement do not necessarily result in lower false alarm rates compared to standard yes/no or old/new recognition tasks (Hicks & Marsh, 1999, 2001). It is possible that when source decisions are required and the source responses in a test include the correct option, more careful inspection of a memory occurs and results in lower false alarm rates. This occurs regardless of whether or not individuals are specifically instructed to examine their memories more carefully.

The previous discussions highlight an important question: why do people not source-monitor or not source-monitor as effectively, when not instructed to? One explanation put forward by Tun et al. (1998) is that participants of any age will employ the most effortless strategy when they are not aware that false memories are being examined. Because the most effortless strategy would appear to be effective, little consideration is given to identifying the source of items, and recognition decisions rely heavily on gist traces, a sense of familiarity, or the activation of the item. This may correspond to participants being less conservative with their responding and less careful in how they inspect their memories compared to when they are instructed to source-monitor. This seems to have been the case in the current research. Similar to Tun et al.'s explanation, it may be that people do consider their memory to be accurate across various situations or tasks. If people genuinely believe their memories are accurate they would be unlikely to spontaneously employ, or be less likely to rely on, a monitoring process. Hence, recognition decisions would heavily rely on gist traces, familiarity, or activation.

Comparing the current study's signal detection findings to similar research conducted by Waldie and See (2003) shows that both demonstrate that older adults' higher false alarm rates compared to younger adults' is a result of older adults having poorer sensitivity, as no age group differences were found for response bias. That is, higher false alarms by older adults are related to their impaired ability to differentiate old and new items, not due to less cautious responding. These results show that older adults did not attempt to compensate for their impaired recollection abilities by employing a more liberal (Howard et al., 2006; Suengas et al., 2010) or more conservative (Botwinick, 1984; Danziger, 1980; Okun, 1976) criterion than younger adults, as is thought to be the case by some.

In line with other research (e.g., Benjamin, 2001; Hicks & Marsh, 2001) sensitivity and/or response bias group differences cannot be used to simply explain group differences in the basic accuracy analyses for false alarm and hit rates. For instance, for overall false alarms both age groups had similar bias scores, but older adults were less sensitive than younger adults. Despite this age-related difference in sensitivity, no significant age group differences in recognition were found for the overall false alarm measure or presented items. When signal detection indices were calculated using critical lure false alarms only, the only age-related difference was that younger adults in the source-monitoring condition had better sensitivity than the other three groups. This differed from the basic accuracy analysis of critical lure false alarms which showed a difference between age groups and test conditions. By contrast, the weak lure analyses found no age group differences, while those in the source-monitoring condition falsely recognised fewer weak lures and had better sensitivity than those in the standard condition. These results suggest that for weak lures, the differences in false recognition between the test conditions can be explained by differences in sensitivity. This is possibly due to those in the source-monitoring condition inspecting their memory of an item more meticulously. It is possible that the differences between the analyses for critical lures and the overall false alarm measure are because some of the effects were too small to be detected with group sizes of 20; thus, further research with greater statistical power is called for (this is discussed further in the limitations).

Overall, the signal detection results were useful for analysing the data and providing further insight into the differences between age groups and test conditions. The results highlighted differences between the groups that the more basic analyses of true and false recognition rates (hit and false alarm rates used outside the signal detection framework) alone could not provide. For instance, for the overall false alarm rate no age-related difference was found, but an age-related difference for sensitivity was found. If only analyses on recognition accuracy had been performed this difference would have been missed. Because signal detection results may find group differences whilst recognition accuracy analyses do not and vice versa, it is useful to conduct both types of analyses to gain a thorough representation of a study's findings. The sensitivity and bias findings have been helpful in providing further support for explanations discussed under the activation-monitoring and fuzzy-trace theories. However, it seems sensitivity and bias differences alone cannot explain the age and condition group differences in false recognition found in the present research. Sensitivity and/or response bias differences between groups do not necessarily result in group differences of false alarm or hit rates. To explain the recognition results sensitivity and/or bias differences would need to be reflected in recognition differences, which for the age group differences in overall false alarms and critical lures they do not.

Implications and Applications

The previous sections have discussed the findings of the present research and how they relate to false memory research and the theories that explain the creation of false memories. The present section moves away from theoretical considerations to focus on ways in which the present research's findings apply to real life and everyday false memories.

As discussed earlier false, repressed, recovered, and discovered memories are theoretically related (Conway, 1997). Yet, little empirical research exists to substantiate this theoretical connection (Kihlstrom, 2004). It is believed false memory research conducted in a laboratory setting is unlikely to mimic the complexities of recovered memories (M. K. Johnson, Raye, Mitchell, & Ankudowich, 2012; Roediger & McDermott, 1996). Caution is warranted when generalising research findings to recovered memories (Freyd & Gleaves, 1996; Gleaves, Smith, Butler, & Spiegel, 2004; Pezdek & Lam, 2007; Roediger & McDermott, 1996). Despite this caution many researchers argue that laboratory research into false memories can generalise and be relevant to other real life situations (M. K. Johnson et al., 2012; Lindsay & Read, 1994; Wade et al., 2007) and a wider set of stimuli (Gallo, 2006; Roediger & McDermott, 1996). Gallo (2006) states that continuity must exist in the underlying cognitive processes or mechanisms that cause false memories, and that false memories are likely a side-effect of information processing mechanisms that are generally useful (e.g., activating associated information, or simplifying information into a gist representation).

A problem with generalising findings from one false memory task to another and to more complex situations is that experimental findings are often dependent on the specific factors of a task (Gallo, 2006). One such example has been discussed above; the difference in results between the present research and Tun et al.'s (1998) study which appear to be due to task differences. Nonetheless, it is widely agreed that at a broader level false memory research is generalisable (e.g., Gallo, 2006; M. K. Johnson et al., 2012; Roediger & McDermott, 1996; Wade et al., 2007). In general terms the present research adds to the evidence that older adults are sometimes, but not always, more prone to false memory errors compared to younger adults. Also, source-monitoring and using a more conservative criterion both appear to play a role in avoiding false memories. These factors suggest that in real life situations it may help to try to recall the context of a memory so a more accurate judgement or decision can be made in an attempt to avoid memory errors. This would be especially relevant in situations where it would be pertinent not to make memory errors, such as eye witness testimonies, or trying to remember if you turned the oven off before leaving the house or only imagining doing it (these situations are discussed further below).

Another broad implication of the present research is the ease with which false memories can be elicited from both younger and older adults, even under conditions in which individuals are required to monitor and respond with the source of their memory. This implies that in at least some everyday situations memories may be false, and the subjective experience of remembering should not be taken as evidence that the event occurred (Roediger & McDermott, 1995; Zaragoza & Lane, 1994), even if a person believes they are correctly remembering the source of the memory. In fact, Roediger and McDermott (1995) believe that false memories may arise more easily in everyday settings, because remembering in real life contexts probably involves a greater level of construction than remembering in experimental tasks (Roediger & McDermott, 1996). It is the constructive nature of remembering that can cause errors (Bartlett, 1932). Gallo (2006) also believes that the ease in which false memories can be elicited suggests that many real life memories may be false, because from an evolutionary perspective our past environments may not have been so rich in information as they are today. Consequently, the likelihood of creating false memories due to associations and similarities between information may be greater now than it has been in our past (Gallo, 2006).

One area in which false memories can have detrimental results is in forensic science (Brainerd & Reyna, 2005). The ease with which false memories can be created by associative or suggestive information increases concerns regarding the use of practices that involve using such information in real life settings (Brainerd & Reyna, 2005; M. K. Johnson et al., 2012). Brainerd and Reyna (2005) have outlined a number of interview techniques that are leading or suggestive and may cause false remembering; for example, yes/no, multiple choice, and repeating questions. These techniques are especially concerning in situations where witnesses are initially questioned, then are interviewed again at a later point, because individuals may incorrectly attribute the source of information from the interview to the event. Even though older adults tend to be disproportionately influenced by suggested information compared to younger adults, both age groups can make these types of errors (G. Cohen & Faulkner, 1989; Dodson & Krueger, 2006; Loftus et al., 1992). The findings of the current research imply that older adults may be more susceptible to these errors compared to younger adults when the false information is strongly associated with the event. An example of a real life situation could be the witnessing of a hit and run incident involving a silver truck. If in the time between the event and making a statement the witness saw a number of similarly coloured trucks (e.g., white or grey), an older adult may well be more likely than a younger adult to falsely remember the truck in the incident being white or grey instead of silver. To decrease the occurrence of false memories when a witness is recalling an event it may help to encourage them to recall as much source information for the event as possible, as well as to avoid leading or suggestive questions.

Other real life false memories, which are more common than criminal or legal situations where eyewitness testimony is important, are often caused by source-monitoring failures (Burke & Light, 1981; Gallo, 2006; Mammarella, Fairfield, & Cornoldi, 2007; Roediger & McDermott, 1996). Everyday false memories can occur when one has difficulty reality monitoring; such as, discriminating real from imagined events or actions from the intent to perform the actions (M. K. Johnson et al., 1993). An implication of the present study's findings is that everyday false memories occur more often for older than younger adults when the true and false memories are strongly related. Such is the case in the reality monitoring paradigm, and older adults are often found to perform more poorly than younger adults (G. Cohen & Faulkner, 1989; Hashtroudi et al., 1989; Henkel et al., 1998; McDaniel et al., 2008; Rosa & Gutchess, 2011). Another everyday false memory can occur when one has difficulty discriminating between sources of facts, especially reliable and unreliable sources. The source recollected for a fact provides valuable information to evaluate the fact's veracity (M. K. Johnson et al., 1993). If one were to remember safety information about a product being from a reputable source instead of from a non-expert friend or tabloid magazine serious consequences could result. To illustrate, incorrectly believing an over-the-counter medication/supplement can be safely taken alongside any other medications/supplements, or incorrectly thinking a heater can be safely left on while one is sleeping may lead to serious harm. Although the present research did not use external sources, previous research has found older adults have more difficulty than younger adults recollecting which of two external sources presented information (Ferguson et al., 1992; McIntyre & Craik, 1987; Schacter et al., 1991). Therefore, older adults may be more prone to these everyday errors than younger adults.

An implication for future research is the importance of attempting to identify conditions that may decrease false memories. In the present research the source-monitoring test included the correct source(s) of false items. Those in the source-monitoring condition had lower false recognition rates than those in the standard condition. The implication of this finding is that the use of source-monitoring in everyday experience may be altered based on how a question is asked. That is, including a source in the question (e.g., did you read that in a *tabloid magazine*? Did *Julie* or *Sam* tell you that?) may make a person check their memory for characteristics specific to the source(s) mentioned. This would likely be more efficient than considering all potential sources (R. L. Marsh & Hicks, 1998). However, it is possible that this technique could later cause false memories if the attempted recollection of characteristics of the specific source mentioned in the question interfered with recollecting the correct source. This raises the question of how to get the right balance. This type of questioning approach may be appropriate and useful in some real life or everyday situations, but the possibility of it causing false memories makes it inappropriate for some settings, such as eyewitness situations.

Limitations and Recommendations

The section above has discussed implications of the current research's findings to real life and everyday false memories. Overall, it appears older adults are more prone to real life and everyday false memories, and that recollecting source information may decrease the incidence of these memory errors in both younger and older adults. Before providing the final conclusions of the current research project the following section discusses limitations that have been identified throughout the project and provides recommendations based on these limitations. Limitations and recommendations based on the DRM false memory task are discussed first, followed by more general limitations and recommendations.

False memory task limitations and recommendations. Despite the current research finding a number of interesting results, several limitations have been identified with the DRM task utilised. It is possible that the DRM lists did not adequately discriminate individual/group performance of the false memory task, due to a possible ceiling (or near ceiling) effect for critical lures. The present research has demonstrated that the DRM lists used, and the corresponding recognition test, resulted in medium to high rates of false recognition across the four age and condition groups. This is despite the present research attempting to increase the variation of false recognition of critical lures by using a combination of DRM lists found in previous research to produce low, medium, or high false recognition rates.

To avoid a ceiling effect and better discriminate individual/group performance, future research will need to go beyond simply choosing lists based on their expected (published) recognition rates, as it appears they are not necessarily accurate (or transferable to populations other than those used in the research). For example, if age differences are to be examined, it may be helpful to use lists that have been employed in previous research in which significant age-related differences were obtained. Also, examining standard deviations associated with DRM lists would likely help identify ones useful for finding individual differences, which may be useful for finding group differences if the variation is explained by group membership. The downside to using previous research to choose DRM lists is that the reliability of the findings and how the findings translate to groups with different characteristics (e.g., from different countries or demographic groups) are unknown.

Another method to avoid ceiling effects may be to increase the number of critical lures presented at test by increasing the number of DRM lists that are studied. Unfortunately, one of the characteristics of DRM lists is that each has 15-items and only a single critical lure. Thus, increasing the number of studied lists makes the study phase substantially longer. A possible alternative method would be to increase the number of critical lures by presenting more lists, but decrease the number of items studied for each list. Lists of as few as three items have been shown to result in moderate false recognition rates (E. Marsh & Bower, 2004), and lists of five (Gallo & Roediger, 2003) or six (Hutchison & Balota, 2005) items have resulted in high false recognition rates. Presenting fewer items allows the opportunity for a greater number of DRM lists to be employed, and is believed to decrease the activation of lure words, resulting in fewer false memories (Robinson & Roediger, 1997). Therefore, the occurrence of a ceiling effect is less likely.

Another concern with the DRM procedure employed was that including weak lures was intended to provide a more sensitive measure of false recognition. In the current study false recognition of the weak lure test items did not follow the findings from previous research (Tun et al., 1998). An age group difference was found for false recognition of critical lures but not weak lures, and the discriminant function analysis showed the item type that was best at differentiating groups was critical lures. These results indicate that weak lures were not sensitive to age group differences for false recognition. Nonetheless, weak lures were sensitive to test group differences; subsequently, the inclusion of these items at test may still be useful in future research. One way to improve their sensitivity might be to use a different method from the present research to select them. For example, if DRM lists shorter than 15 items were presented weak lures could be words from the standard 15-item lists that were not presented. Alternatively, measures other than accuracy could be used. For instance, Tun et al (1998) found age group differences for response latencies which provided valuable information for the theoretical frameworks they discussed.

A final limitation of the DRM task was the length of the test (96 items, each requiring a recognition response and a certainty rating). This was a potential confound because generally, as expected, the test took older adults slightly longer (5 to 10 minutes) to complete than younger adults. In some cases it took older adults much longer to complete. This was related to a small number of older adults finding the task too difficult to finish, as at a point during the test they stopped as they believed they could no longer recognise which items may or may not have been presented. As a result, these individuals were excluded from the final sample. It is clear that for those individuals unable to complete the task the time involved and the cognitive demands were too much. Participants may have benefited from the DRM task being made into two separate and smaller tasks, so a break could have been provided. This may have made it easier for some to complete the task, and then data would not have to be excluded. However, aside from a small number of older adults, no other participants (younger or older) indicated they found the task too difficult. The limitations of the DRM task identified above will be useful to consider for future research.

General limitations and recommendations. One general concern, was the possibility some of the effects were small, thus difficult to detect with the present study's group sizes of 20. Future research planning to examine age-related differences in DRM tasks could profit by attempting to recruit more participants for their study to improve power (Field, 2009). Recruiting participants (in particular older adults) can be quite difficult and time consuming, as was found in the present research. Therefore, depending on the practical scope and time constraints of the research to be undertaken this may not be a viable option. Alternatively, statistical power may be improved by using more precise measurement tools. For example, using DRM lists found to result in greater effect sizes than other lists. Employing more than one false memory task and collating or triangulating the results might decrease measurement error and increase power. Other false memory tasks include the false-fame paradigm (Jacoby et al., 1989) and the eyewitness suggestibility paradigm (Loftus, 1979).

Also possibly related to small effects difficult to detect with group sizes of 20, the current research highlighted an issue when analysing the two components (false alarm and hit rates) of sensitivity (*d'*) and bias (*c*) separately. This is because both the false alarm and hit rate can be affected by a change in sensitivity *and* a change in bias. The benefit of the signal detection framework is that false alarms and hits are combined to produce sensitivity and bias estimates. A problem in the present research was that, for the test conditions, differences were found for both overall sensitivity and overall response bias. Because both differences were found it is unclear which is producing the difference in the overall false alarm rate found between the test conditions. Therefore, this finding must be interpreted with caution.

Another concern is the generalisability of the present research. Bäckman et al. (2000) believe problems exist when generalising age-related research findings when the participants in a study were young-old (55 to 74 years old) to the old-old (75-85 years) and, oldest-old (85 years and older). Because the older adults in the current research were aged 75-80 (old-old), it follows that generalising results to those considered young-old or oldest-old may be problematic. The problem is whether the performance of the older adults would be comparable to those approximately 10 years younger or older, or whether age-related performance on the DRM task would be gradual or accelerated with increasing age. Attempting to discover the answer to this question by comparing groups considered young-old, old-old, and oldest-old would be an interesting topic for future research. It would also provide further insight into the mechanisms behind the creation of false memories.

Despite planning the options to identify people’s education level to include educational opportunities relevant to both age groups, the options worked for younger adults but not older adults. Recording older participants’ level of education was often difficult and tended to require a discussion to see which option would best represent their educational history. Consequently, education levels between the two age groups may not be comparable, and age group differences might be confounded by education level. Instead of using a recording system like the one used in the present study, future research that employs participant groups of different generations with different educational opportunities may benefit from avoiding education and instead measuring a related construct, such as intelligence.

The cut-off score for the MoCA had to be altered (from 26 to 24) in the present research, indicating it may not have been the ideal cognitive measure to use as a screening tool. It was not until all of the younger adults’ and approximately a third of the older adults’ data had been collected that this issue became evident (about a third of the older adults were scoring below 26), making it impractical to change to another measure. Despite all of the individuals included in the study being community dwelling adults who were able to complete the DRM task without difficulty, it is possible that decreasing the cut-off score resulted in some individuals with unidentified cognitive problems being included in the study, which would have led to confounded outcomes. Future research that is going to use a brief cognitive screening tool with a New Zealand sample should consider alternative measures to the MoCA. One viable alternative measure is the Addenbrooke's Cognitive Examination-Revised (ACE-R; Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006). The ACE-R is commonly used in New Zealand to screen for mild cognitive impairment (Strauss, Leathem, Humpries, & Podd, 2012). Like the MoCA the ACE-R has been found to be acceptably reliable with high sensitivity and specificity at detecting mild cognitive impairment (Mioshi et al., 2006).

A difficulty with DRM research is that the two main theories explaining false memory creation overlap. Often both theories can explain the same set of results, making it difficult to identify the exact processes or mechanisms that likely underpin age-related differences. The above discussion is evidence of this difficulty. It appears the only way to attempt to establish which of the two theories is best at explaining false memory creation and age-related differences is to continue studying the DRM paradigm and altering the DRM tasks in ways that might provide the necessary information. For instance, one change that might influence the use of gist traces would be to completely exclude unrelated words from recognition tests. This would mean all items contained in a test would be semantically related, making gist processes much less efficient for making recognition decisions (Gunter, Ivanko, & Bodner, 2005). Recording response latencies may be helpful, as Gallo (2006) believes differences in response latencies between lures and studied words likely reflect differences in the processes involved in the creation of false memories.

Gallo (2004, 2006, 2010) elaborated on the monitoring processes involved in the creation of false memories and identified two different types: disqualifying and diagnostic monitoring. Although currently this idea is under-developed it provides a more sophisticated way of understanding false memories than that given in existing formulations of the fuzzy-trace and activation-monitoring theories. Unfortunately, because the disqualifying and diagnostic monitoring frameworks are currently under-developed it is difficult to identify what type of DRM task, or what manipulations, will be useful for examining one of the monitoring processes over the other (Gallo, 2004, 2010). Future research will clearly benefit from considering the limitations and recommendations discussed above that were exposed during the present research. Additionally, if future research were planning to further examine the signal detection false memory models, the following recommendations would be crucial.

As it was not one of the aims of the present research to examine the two signal detection models of false recognition, the present study was unable to investigate how the data corresponded to each of the models separately. If an aim of future research was to examine these models the research would need to use procedures similar to Miller and Wolford (1999), Miller et al. (2011), and Westerberg and Marsolek (2003), in which recognition of presented and non-presented critical lures, related words, and unrelated words is tested. Using this procedure allows for independent measures of sensitivity and bias to be calculated, providing separate indices for each item type, and thus estimates of each distribution and criterion placement along the memory strength continuum. The different indices can then be compared to establish which of the two models fits the data best.

Conclusions

Several conclusions can be made based on the results of the current study. Firstly, the age-related differences in the DRM false recognition task were not best explained by the fuzzy-trace theory, source-monitoring processes, or criterion and/or sensitivity differences in signal detection ability. As expected, specific results were found that partially favoured either source-monitoring processes (as part of the activation-monitoring theory) or the fuzzy-trace theory. The activation-monitoring theory and the fuzzy-trace theory appear flexible in how they can be used to explain both expected and unexpected results. The majority of results were able to be adequately explained by both theories, partly because of the significant overlap in some of the characteristics of the two theoretical approaches. However, source-monitoring processes as part of the activation-monitoring theory more easily applied to and explained the present research's findings compared to the fuzzy-trace theory or criterion and/or sensitivity differences. Conclusions regarding each of the three possible explanations are further outlined below.

Importantly, one conclusion is that source-monitoring processes appear to play an important role in the creation of false memories. Regardless of age, requiring participants to source-monitor resulted in fewer new items (critical lure, weak lure, and overall false alarms) being incorrectly classified as old, and resulted in participants using a more conservative criterion. It appears that requiring source-monitoring encouraged participants to use item-specific processing; focusing on source-specifying characteristics to make more accurate decisions than in the standard test condition. Source memory is also important to both the fuzzy-trace and activation-monitoring theory. In the activation-monitoring theory source memory falls under source-monitoring; one of the two mechanisms in the theory. In the fuzzy-trace theory source characteristics are considered verbatim traces. An interesting finding from the test manipulation was that requiring source-monitoring disproportionately improved younger adults' sensitivity of critical lures compared to older adults, but did not result in younger adults recognising disproportionately fewer critical lures than older adults. This inconsistency was possibly due to small sample sizes, and it would make an interesting topic for further research.

Activation-monitoring theory explains the findings in the following way: in the standard test condition both age groups employed a source-monitoring strategy and the source-monitoring condition enhanced this already-used strategy. As expected, older adults' source-monitoring strategy was impaired. Older adults falsely recognised more critical lures and were more certain falsely recognised critical lures and weak lures had been heard compared to younger adults. Additionally, based on the activation-monitoring theory some results were unexpected, but were able to be adequately explained. Firstly, no age group differences were found for false recognition of weak lures. Although unexpected this finding can be explained by the possibility both age groups' source-monitoring ability was effective at identifying weak lures as new, possibly due to available differences in item-specific characteristics and semantic variation. Secondly, no age group differences were found for recognition of presented items. Again, unexpected based on older adults memory impairments due to cognitive aging, but the finding implies that the processes or strategies used for false memory tasks are different from general memory tasks, and that both effortful and automatic processes are employed in DRM false memory tasks. Thirdly, it was expected that older adults performance would improve disproportionately compared to younger adults when source-monitoring was required, because it was believed that in the standard condition older adults would not use a source-monitoring strategy. Yet, both age groups performance improved equally. This finding supports the idea that both age groups employed a source-monitoring strategy in the standard test, and requiring source-monitoring enhanced the strategy that was already being employed.

For the fuzzy-trace theory the following results are explained well: (a) older adults falsely recognised more critical lures than younger adults, because older adults rely more on gist traces than younger adults; (b) recognition of presented items was similar across the two age groups, because older adults reliance on gist traces can compensate for their deficit in recollecting verbatim traces; and (c) older were more certain than younger adults that critical lures, weak lures, and presented items were heard, because these items would have strong gist representations if participants relied on gist traces they would be more certain than if they relied on both gist and verbatim traces. The results that are not adequately explained by the fuzzy-trace theory are those from the source-monitoring condition. Requiring source-monitoring was expected to improve both age groups performance; but, older adults' improvement should not have equalled that of younger adults, due to their impaired verbatim processes. This result indicates older adults were able to use verbatim traces to offset gist traces as well as younger adults. Lastly, the fuzzy-trace theory cannot explain why no age group differences were found for false recognition of weak lures, or overall false alarms. Older adults' reliance on gist traces more than younger adults should have resulted in higher rates of false recognition for all measures.

Another conclusion is that the signal detection results provided valuable information towards explaining group differences in the creation of false memories. The bias results demonstrated that those in the source-monitoring condition responded more cautiously than those in the standard condition, even though they were not specifically instructed to do so. Yet, their cautious responding did not result in fewer hits than the standard condition, suggesting they were able to adopt a criterion that maximised hits while minimising false alarms. Those in the source-monitoring condition also had better sensitivity than those in the standard condition, supporting the idea that requiring a source judgement (when the correct source option is in the recognition test) made participants inspect their memories more meticulously.

Again, the importance of source-monitoring in false memory tasks was highlighted by the signal detection results. It is possible that when not instructed to, people do not spontaneously use a source-monitoring strategy, or use it as effectively as they would when they are instructed to. This maybe because they employ the most effortless strategy or believe their memories are in fact accurate; hence, they rely more heavily on gist traces, familiarity, and activation. The signal detection results also demonstrated that older adults did not attempt to compensate for their impaired recollection abilities by using less cautious responding compared to younger adults. However, bias and/or sensitivity differences alone cannot fully explain age and condition group differences in false recognition. Sensitivity and bias differences were found that were not reflected in group differences for false recognition, possibly due to some group differences being too small to be detected. If one wanted to further investigate the role of sensitivity and criterion in the creation of false memories it would be useful to employ procedures that allow further examination of the signal detection models of false recognition.

Finally, the results of the present study highlight the robustness of DRM false memories; under both a standard condition and a condition that required source-monitoring false memories of critical lures were created, and participants were highly certain that the critical lures they falsely recognised were heard during testing when they were not. Furthermore, older adults were more prone to false memories than younger adults. These findings have implications for how real life and everyday false memories might in fact be a common occurrence. Real life false memories can arise in eyewitness situations, with problems differentiating real from imagined events and actions from plans to execute the action, and mistakes in the veracity of facts based on the recollected source of the information. Also, false recognition occurred at a lower rate in the source-monitoring condition than in the standard condition. This finding implies that having one recollect more context information than they generally might, could decrease the incidence of false memories in both younger and older adults.

References

Aguinis, H., Gottferdson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods, 16*, 270-301. doi:10.1177/1094428112470848

Alberts, J. (2007). When our memories fail us: Exploring the accuracy and inaccuracy of memory. *New Zealand Clinical Psychology Journal, 17*, 15-21.

An, Q., Xu, D., & Brooks, G. P. (2013). Type I error rates and power of multiple hypothesis testing procedures in factorial ANOVA. *Multiple Linear Regression Viewpoints, 39*, 1-16. Retrieved from http://mlrv.ua.edu/2013/vol39_2/An-etal-39_2_proof_1.pdf

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language, 39*, 371-391. doi:10.1006/jmla.1998.2581

Bäckman, L., & Larsson, M. (1992). Recall of organizable words and objects in adulthood: Influences of instructions, retention interval, and retrieval cues. *Journal of Gerontology, 47*, 273-278. doi:10.1093/geronj/47.4.P273

Bäckman, L., Small, B. J., Wahlin, Å., & Larsson, M. (2000). Cognitive functioning in very old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 499-558). Mahwah, NJ: Erlbaum.

Baddeley, A. (1986). *Working memory*. Oxford, England: Clarendon Press.

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., III, McDermott, K. B., et al. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology, 16*, 361-384. doi:10.1080/026432999380834

Barnett, V., & Lewis, T. (1978). *Outliers in statistical data*. Chichester, England: Wiley & Sons.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). Chichester, England: Wiley & Sons.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 941-947. doi:10.1037/0278-7393.27.4.941

Binet, A. (1900). *La suggestibilité*. Paris, France: Schleicher Freres.

Borella, E., Carretti, B., & De Beni, R. (2008). Working memory and inhibition across the adult life-span. *Acta Psychologica, 128*, 33-44. doi:10.1016/j.actpsy.2007.09.008

Botwinick, J. (1984). *Aging and behavior: A comprehensive integration of research findings* (3rd ed.). New York, NY: Springer.

Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experiences are easier to "remember" than things that were. *Psychological Science, 9*, 484-489. doi:10.1111/1467-9280.00089

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

Brainerd, C. J., & Reyna, V. F. (2002b). Recollection rejection: How children edit their false memories. *Developmental Psychology, 38*, 156. doi:10.1037/0012-1649.38.1.156

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. Oxford, NY: Oxford University Press.

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language, 34*, 157-185. doi:10.1006/jmla.1995.1008

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review, 110*. doi:10.1037/0033-295X.110.4.762

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*. doi:10.1037/0278-7393.27.2.307

Bray, J. H., & Maxwell, S. E. (1982). Analyzing and interpreting significant MANOVAs. *Review of Educational Research, 52*, 340-367. doi:10.3102/00346543052003340

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance: Vol. 54. Sage university paper series on quantitative research methods*. Newbury Park, CA: Sage.

Brueckner, K., & Moritz, S. (2009). Emotional valence and semantic relatedness differentially influence false recognition in mild cognitive impairment, Alzheimer's disease, and healthy elderly. *Journal of the International Neuropsychological Society, 15*, 268-276. doi:10.1017/S135561770909047X

Budson, A. E., Daffner, K. R., Desikan, R., & Schacter, D. L. (2000). When false recognition is unopposed by true recognition: Gist-based memory distortion in Alzheimer's disease. *Neuropsychology, 14*, 277-287. doi:10.1037//0894-4105.14.2.277

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition, 51*, 251-261. doi:10.1016/S0278-2626(03)00030-7

Burke, D. M., & Light, L. L. (1981). Memory and aging: The role of retrieval processes. *Psychological Bulletin, 90*, 513-546. doi:10.1037/0033-2909.90.3.513

Butler, K. M., McDaniel, M. A., Dornburg, C. C., Price, A. L., & Roediger, H. L., III. (2004). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin and Review, 11*, 921-925. doi:10.3758/BF03196722

Cardinal, R. N., & Aitken, M. R. F. (2006). *ANOVA for the behavioural sciences researcher*. Mahwah, NJ: Erlbaum.

Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General, 132*, 310-324. doi:10.1037/0096-3445.132.2.310

Cherry, K. E., & Park, D. C. (1993). Individual difference and contextual variables influence spatial memory in younger and older adults. *Psychology and Aging, 8*, 517-526. doi:10.1037/0882-7974.8.4.517

Cohen, G., & Faulkner, D. (1989). Age differences in source forgetting: Effects on reality monitoring and on eyewitness testimony. *Psychology and Aging, 4*, 10-17. doi:10.1037/0882-7974.4.1.10

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1994). How the power of MANOVA can both increase and decrease as a function of the intercorrelations among the dependent variables. *Psychological Bulletin, 115*, 465-474. doi:10.1037/0033-2909.115.3.465

Colsher, P. L., & Wallace, R. B. (1991). Longitudinal application of cognitive function measures in a defined population of community-dwelling elders. *Annals of Epidemiology, 1*, 215-230. doi:10.1016/1047-2797(91)90001-S

Conway, M. A. (1997). Past and present: Recovered memories and false memories. In M. A. Conway (Ed.), *Recovered memories and false memories* (pp. 150-191). Oxford, NY: Oxford University Press.

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191-211). New York, NY: Plenum Press.

Craik, F. I. M., & Jennings, J. M. (1992). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 51-110). Hillsdale, NJ: Erlbaum.

Crook, T. H., & Larrabee, G. J. (1992). Changes in facial recognition memory across the adult life span. *Journal of Gerontology, 47*, 138-141. doi:10.1093/geronj/47.3.P138

Crook, T. H., & West, R. L. (1990). Name recall performance across the adult life-span. *British Journal of Psychology*, 335-349. doi:10.1111/j.2044-8295.1990.tb02365.x

Curran, T., Schacter, D. L., Norman, K. A., & Galluccio, L. (1997). False recognition after a right frontal lobe infarction: Memory for general and specific information. *Neuropsychologia, 35*, 1035-1049. doi:10.1016/s0028-3932(97)00029-8

Danziger, W. L. (1980). Measurement of response bias in aging research. In L. W. Poon (Ed.), *Aging in the 1980s: Psychological issues* (pp. 552-557). Washington, DC: American Psychological Association.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17-22. doi:10.1037/h0046671

Dehon, H., & Brédart, S. (2004). False memories: Young and older adults think of semantic associates at the same rate, but young adults are more successful at source monitoring. *Psychology and Aging, 19*, 191-197. doi:10.1037/0882-7974.19.1.191

Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging, 4*, 500-503. doi:10.1037/0882-7974.4.4.500

Dodson, C. S., Bawa, S., & Slotnick, S. D. (2007). Aging, source memory, and misrecollections. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 169-181. doi:10.1037/0278-7393.33.1.169

Dodson, C. S., & Krueger, L. E. (2006). I misremember it well: Why older adults are unreliable eyewitnesses. *Psychonomic Bulletin and Review, 13*, 770-775. doi:10.3758/BF03193995

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 and Review, 8*, 155-161. doi:10.3758/BF03196152

Dodson, C. S., & Schacter, D. L. (2002). Aging and strategic retrieval processes: Reducing false memories with a distinctiveness heuristic. *Psychology and Aging, 17*, 405-415. doi:10.1037//0882-7974.17.3.405

Duro, D., Simões, M. R., & Ponciano, E. (2010). Validation studies of the Portuguese experimental version of the Montreal Cognitive Assessment (MoCA): Confirmatory factor analysis. *Journal of Neurology, 257*, 728-734. doi:10.1007/s00415-009-5399-5

Dywan, J., & Jacoby, L. (1990). Effects of aging on source monitoring: Differences in susceptibility to false fame. *Psychology and Aging, 5*, 379-387. doi:10.1037/0882-7974.5.3.379

Ebbinghaus, H. (1885/1964). *Memory: A contribution to experimental psychology.* New York, NY: Dover.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309-331. doi:10.1037/0096-3445.128.3.309

Erdelyi, M. H. (1985). *Psychoanalysis: Freud's cognitive psychology*. New York, NY: Freeman.

Fairfield, B., & Mammarella, N. (2009). The role of cognitive operations in reality monitoring: A study with healthy older adults and Alzheimer's-type dementia. *Journal of General Psychology, 136*, 21-39. doi:10.3200/GENP.136.1.21-40

Farcomeni, A. (2007). A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Statistical Methods in Medical Research, 17*, 347-388. doi:10.1177/0962280206079046

Ferguson, S. A., Hashtroudi, S., & Johnson, M. K. (1992). Age differences in using source-relevant cues. *Psychology and Aging, 7*, 443-443. doi:10.1037/0882-7974.7.3.443

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, England: Sage.

Freud, S. (1896). The aetiology of hysteria. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud.* London, England: Hogarth.

Freyd, J. J., & Gleaves, D. H. (1996). "Remembering" words not presented in lists: Relevance to the current recovered/false memory controversy. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 22, 811-813. doi:10.1037/0278-7393.22.3.811

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 120-128. doi:10.1037/0278-7393.30.1.120

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York, NY: Psychology Press.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38*, 833-848. doi:10.3758/MC.38.7.833

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory, 14*, 730-741. doi:10.1080/09658210600648506

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin and Review, 4*, 271-276. doi:10.3758/BF03209405

Gallo, D. A., & Roediger, H. L., III. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47*, 469-497. doi:10.1016/S0749-596X(02)00013-X

Gallo, D. A., & Roediger, H. L., III. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition, 31*, 1036-1044. doi:10.3758/BF03196124

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*, 83-98. doi:10.1080/09658210143000281

Ge, Y., Sealfon, S. C., Tseng, C., & Speed, T. P. (2007). A Holm-type procedure controlling the false discovery rate. *Statistics & Probability Letters, 77*, 1756-1762. doi:10.1016/j.spl.2007.04.019

Gerrie, M. P., & Garry, M. (2007). Individual differences in working memory capacity affect false memories for missing aspects of events. *Memory, 15*, 561-571. doi:10.1080/09658210701391634

Gick, M. L., Craik, F. I. M., & Morris, R. G. (1988). Task complexity and age differences in working memory. *Memory & Cognition, 16*, 353-361. doi:10.3758/BF03197046

Gilinsky, A. S., & Judd, B. B. (1994). Working memory and bias in reasoning across the adult life span. *Psychology and Aging, 9*, 356-371. doi:10.1037/0882-7974.9.3.356

Gill, D. J., Freshman, A., Blender, J. A., & Ravina, B. (2008). The Montreal Cognitive Assessment as a screening tool for cognitive impairment in Parkinson's disease. *Movement Disorders, 23*, 1043-1046. doi:10.1002/mds.22017

Ginsburg, N., & Karpiuk, P. (1994). Random number generation: Analysis of responses. *Perceptual and Motor Skills, 79*, 1059-1067. doi:10.2466/pms.1994.79.3.1059

Gleaves, D. H., Smith, S. M., Butler, L. D., & Spiegel, D. (2004). False and recovered memories in the laboratory and clinic: A review of experimental and clinical evidence. *Clinical Psychology: Science and Practice, 11*, 3-28. doi:10.1093/clipsy.bph055

Goldstein, E. B. (2010). *Sensation and perception* (8th ed.). Belmont, CA: Wadsworth.

Gordon, S. K., & Clark, W. C. (1974a). Adult age differences in word and nonsense syllable recognition memory and response criterion. *Journal of Gerontology, 29*. doi:10.1093/geronj/29.6.659

Gordon, S. K., & Clark, W. C. (1974b). Application of signal detection theory to prose recall and recognition in elderly and young adults. *Journal of Gerontology, 29*. doi:10.1093/geronj/29.1.64

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

Gunter, R., Ivanko, S., & Bodner, G. (2005). Can test list context manipulations improve recognition accuracy in the DRM paradigm? *Memory, 13*, 862-873. doi:10.1080/09658210444000458

Hair, J. F., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis: A global perspective* (7th ed.). Harlow, NJ: Pearson Education.

Hancock, T. W., Hicks, J. L., Marsh, R. L., & Ritschel, L. (2003). Measuring the activation level of critical lures in the Deese-Roediger-McDermott paradigm. *American Journal of Psychology, 116*, 1-14. Retrieved from http://www.jstor.org/stable/1423332 .

Harkins, S. W., Chapman, C. R., & Eisdorfer, C. (1979). Memory loss and response bias in senescence. *Journal of Gerontology, 34*, 66-72. doi:10.1093/geronj/34.1.66

Hartley, J. T. (1986). Reader and text variables as determinants of discourse memory in adulthood. *Psychology and Aging, 2*, 150-158. doi:10.1037/0882-7974.1.2.150

Hartley, J. T. (1993). Aging and prose memory: Tests of the resource-deficit hypothesis. *Psychology and Aging, 8*, 538-551. doi:10.1037/0882-7974.8.4.538

Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 163-169. doi:10.1037/0278-7393.17.1.163

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 193-225). San Diego, CA: Academic Press.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging, 4*, 106-112. doi:10.1037/0882-7974.4.1.106

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and Aging, 5*, 119-126. doi:10.1037/0882-7974.5.1.119

Heaps, C. M., & Nash, M. (2001). Comparing recollective experience in true and false autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 920. doi:10.1037//0278-7393.27.4.920

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*, 251-268. doi:10.1037/0096-3445.127.3.251

Hess, T. M. (1984). Effects of semantically related and unrelated contexts on recognition memory of different-aged adults. *Journal of Gerontology, 39*, 444-451. doi:10.1093/geronj/39.4.444

Hicks, J. L., & Marsh, R. L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1195-1209. doi:10.1037/0278-7393.25.5.1195

Hicks, J. L., & Marsh, R. L. (2001). False recognition occurs more frequently during source identification than during old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 375-383. doi:10.1037/0278-7393.27.2.375

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33*, 1-18. doi:10.1006/jmla.1994.1001

Holland, B. S., & Copenhaver, M. D. (1987). An improved sequentially rejective Bonferroni test procedure. *Biometrics, 43*, 417-423. Retrieved from http://www.jstor.org/stable/2531823

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65-70. Retrieved from http://www.jstor.org/stable/4615733

Hoops, S., Nazem, S., Siderowf, A. D., Duda, J. E., Xie, S. X., Stern, M. B., et al. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology, 73*, 1738-1745. doi:10.1212/WNL.0b013e3181c34b47

Howard, M. W., Bessette-Symons, B., Zhang, Y., & Hoyer, W. J. (2006). Aging selectively impairs recollection in recognition memory for pictures: Evidence from modeling and receiver operating characteristic curves. *Psychology and Aging, 21*, 96-106. doi:10.1037/0882-7974.21.1.96

Hultsch, D. F., Hertzog, C., Small, B. J., McDonald-Miszczak, L., & Dixon, R. A. (1992). Short-term longitudinal change in cognitive performance in later life. *Psychology and Aging, 7*, 571-584. doi:10.1037/0882-7974.7.4.571

Hultsch, D. F., Masson, M. E. J., & Small, B. J. (1991). Adult age differences in direct and indirect tests of memory. *Journal of Gerontology*, 22-30. doi:10.1093/geronj/46.1.P22

Hutchison, K. A., & Balota, D. A. (2005). Decoupling semantic and associative information in false memories: Explorations with semantically ambiguous and unambiguous critical lures. *Journal of Memory and Language, 52*, 1-28. doi:10.1016/j.jml.2004.08.003

Intons-Peterson, M. J., & Best, D. L. (1998). Introduction and a brief history of memory distortions and their prevention. In M. J. Intons-Peterson & D. L. Best (Eds.), *Memory distortions and their prevention* (pp. 1-14). Mahwah, NJ: Erlbaum.

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1999). Age, testing at preferred or nonpreferred times (testing optimality), and false memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 23-40. doi:10.1037/0278-7393.25.1.23

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 3-22. doi:10.1037/0278-7393.25.1.3

Jacoby, L. L., Jennings, J. M., & Hay, J. F. (1996). Dissociating automatic and consciously controlled processes: Implications for diagnosis and rehabilitation of memory deficits. In D. Herrmann, C. McEvoy, C. Hertzof, P. Hertel, & M. K. Johnson (Eds.), *Basic and applied memory research: Theory in context* (Vol. 1, pp. 161-193). Mahwah, NJ: Erlbaum.

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology, 56*, 326-338. doi:10.1037/0022-3514.56.3.326

Jenkins, J. J. (1970). The 1952 Minnesota word association norms. In L. Postman & G. Keppel (Eds.), *Norms of word association* (pp. 1-38). New York, NY: Academic Press.

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging, 8*, 283-293. doi:10.1037/0882-7974.8.2.283

Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition. *Psychology and Aging, 12*, 352-361. doi:10.1037/0882-7974.12.2.352

Johnson, M. K., Bransford, J. D., & Solomon, S. K. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology, 98*, 203-205. doi:10.1037/h0034290

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General, 117*, 371-376. doi:10.1037/0096-3445.117.4.371

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

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85. doi:10.1037/0033-295X.88.1.67

Johnson, M. K., Raye, C. L., Mitchell, K. J., & Ankudowich, E. (2012). The cognitive neuroscience of true and false memories. In R. F. Belli (Ed.), *True and false recovered memories: Toward a reconciliation of the debate* (Vol. 58, pp. 15-52). New York, NY: Springer.

Johnson, R. E. (2003). Aging and the remembering of text. *Developmental Review, 23*, 261-346. doi:10.1016/S0273-2297(03)00009-1

Jou, J. (2011). Conscious and unconscious discriminations between true and false memories. *Consciousness and Cognition, 20*, 828-839. doi:10.1016/j.concog.2010.10.022

Kausler, D. H. (1994). *Learning and memory in normal aging*. London, England: Academic Press.

Kausler, D. H., & Puckett, J. M. (1980). Adult age differences in recognition memory for a nonsemantic attribute. *Experimental Aging Research, 6*, 349-355. doi:10.1080/03610738008258370

Kausler, D. H., & Puckett, J. M. (1981). Adult age differences in memory for modality attributes. *Experimental Aging Research, 7*, 117-125. doi:10.1080/03610738108259794

Kausler, D. H., & Wiley, J. G. (1990). Temporal memory and content memory for actions: Adult age differences in acquisition and retention. *Experimental Aging Research, 16*, 147-150. doi:10.1037/0882-7974.10.4.527

Kensinger, E. A., O'Brien, J. L., Swanberg, K., Garoff-Eaton, R. J., & Schacter, D. L. (2007). The effects of emotional content on reality-monitoring performance in young and older adults. *Psychology and Aging, 22*, 752-764. doi:10.1037/0882-7974.22.4.752

Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of ageing. *Cognitive Neuropsychology, 16*, 399-416. doi:10.1080/026432999380852

Keselman, H. J., Miller, C. W., & Holland, B. (2011). Many tests of significance: New methods for controlling Type I errors. *Psychological Methods, 17*, 420-431. doi:10.1037/a0025810

Kihlstrom, J. F. (2004). An unbalanced balancing act: Blocked, recovered, and false memories in the laboratory and clinic. *Clinical Psychology: Science and Practice, 11*, 34-41. doi:10.1093/clipsy.bph057

Korten, A. E., Henderson, A. A., Christensen, H., Jorm, A. F., Rodgers, B., Jacomb, P., et al. (1997). A prospective study of cognitive function in the elderly. *Psychological Medicine, 27*, 919-930. doi:10.1017/S0033291797005217

Koski, L., Xie, H., & Finch, L. (2009). Measuring cognition in a geriatric outpatient clinic: Rasch analysis of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology, 22*, 151-160. doi:10.1177/0891988709332944

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37*, 555-583. doi:10.1006/jmla.1997.2529

Lamont, A. (2006). *Cross-sectional and longitudinal analyses of the effects of aging on memory in healthy young, middle-aged, and oldest-old adults* (Doctoral thesis, Massey University, Palmerston North, New Zealand). Retrieved from http://hdl.handle.net/10179/258

Lamont, A., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: Effects of target age and memory load. *Memory & Cognition, 33*, 1017-1024. doi:10.3758/bf03193209

Lampinen, J. M., Meier, C. R., Arnal, J. D., & Leding, J. K. (2005). Compelling untruths: Content borrowing and vivid false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 954-963. doi:10.1037/0278-7393.31.5.954

Law, S., Hawkins, S. A., & Craik, F. I. M. (1998). Repetition-induced belief in the elderly: Rehabilitating age-related memory deficits. *Journal of Consumer Research, 25*, 91-107. doi:10.1086/209529

Le Breck, D. B., & Baron, A. (1987). Age and practice effects in continuous recognition memory. *Journal of Gerontology, 42*, 89-91. doi:10.1093/geronj/42.1.89

Leding, J. K. (2012). Working memory predicts the rejection of false memories. *Memory, 20*, 217-223. doi:10.1080/09658211.2011.653373

Lee, J.-Y., Lee, D. W., Cho, S.-J., Na, D. L., Jeon, H. J., Kim, S.-K., et al. (2008). Brief screening for mild cognitive impairment in elderly outpatient clinic: Validation of the Korean version of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology, 21*, 104-110. doi:10.1177/0891988708316855

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42*, 333-376. Retrieved from http://www.annualreviews.org.ezproxy.massey.ac.nz/doi/pdf/10.1146/annurev.ps.42.020191.002001

Light, L. L., & Zelinski, E. M. (1983). Memory for spatial information in young and old adults. *Developmental Psychology, 19*, 901-906. doi:10.1037/0012-1649.19.6.901

Lindsay, D. S., Gonzalez, V., & Eso, K. Z. (1995). Aware and unaware uses of memoires of postevent suggestions. In M. S. Zaragoza, J. R. Graham, G. C. N. Hall, R. Hirschman, & Y. S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (pp. 86-108). Thousand Oaks, CA: Sage.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17*, 349-358. doi:10.3758/BF03198473

Lindsay, D. S., & Johnson, M. K. (2000). False memories and the source monitoring framework: Reply to Reyna and Lloyd (1997). *Learning and Individual Differences, 12*, 145-161. doi:10.1016/s1041-6080(01)00035-8

Lindsay, D. S., & Read, J. D. (1994). Psychotherapy and memories of childhood sexual abuse: A cognitive perspective. *Applied Cognitive Psychology, 8*, 281-338. doi:10.1002/acp.2350080403

Loftus, E. F. (1979). *Eyewitness testimony*. London, England: Harvard University Press.

Loftus, E. F., Levidow, B., & Duensing, S. (1992). Who remembers best? Individual differences in memory for events that occured in a science museum. *Applied Cognitive Psychology, 6*, 93-107. doi:10.1002/acp.2350060202

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19-31. doi:10.1037/0278-7393.4.1.19

Lövdén, M. (2003). The episodic memory and inhibition accounts of age-related increases in false memories: A consistency check. *Journal of Memory and Language, 49*, 268-283. doi:10.1016/S0749-596X(03)00069-X

Luis, C. A., Keegan, A. P., & Mullan, M. (2009). Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the Southeastern US. *International Journal of Geriatric Psychiatry*, 197-201. doi:10.1002/gps.2101

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Mammarella, N., Fairfield, B., & Cornoldi, C. (2007). Reality monitoring and resistance to forgetting under short delay intervals. *The Quarterly Journal of Experimental Psychology, 60*, 551-570. doi:10.1080/17470210600760227

Marsh, E., & Bower, G. (2004). The role of rehearsal and generation in false memory creation. *Memory, 12*, 748-761. doi:10.1080/09658210344000170

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

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition, 25*, 826-837. doi:10.3758/BF03211327

May, C. P., Kane, M. J., & Hasher, L. (1995). Determinants of negative priming. *Psychological Bulletin*, 35-54. doi:10.1037/0033-2909.118.1.35

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition, 30*, 1065-1077. doi:10.3758/BF03194324

McDaniel, M. A., Lyle, K. B., Butler, K. M., & Dornburg, C. C. (2008). Age-related deficits in reality monitoring of action memories. *Psychology and Aging, 23*, 646-656. doi:10.1037/a0013083

McDermott, K. B. (1995). [Associative lists that induce false memories]. Unpublished materials.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35*, 212-230. doi:10.1006/jmla.1996.0012

McDermott, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin and Review, 4*, 582-586. doi:10.3758/BF03214353

McDermott, K. B., & Roediger, H. L., III. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language, 39*, 508-520. doi:10.1006/jmla.1998.2582

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language, 45*, 160-176. doi:10.1006/jmla.2000.2771

McEvoy, C. L., Nelson, D. L., & Komatsu, T. (1999). What's the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1177-1194. doi:10.1037/0278-7393.25.5.1177

McIntyre, J. S., & Craik, F. I. M. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology, 41*, 175-192. doi:10.1037/h0084154

McNicol, D. (1972). *A primer of signal detection theory*. Mahwah, NJ: Erlbaum.

Meyers, L. W., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research: Design and interpretation* (2nd ed.). Thousand Oaks, CA: Sage.

Miller, M. B., Guerin, S. A., & Wolford, G. L. (2011). The strategic nature of false recognition in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1228-1235. doi:10.1037/a0024539

Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review, 102*, 398-405. doi:10.1037/0033-295X.106.2.398

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): A brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry, 21*, 1078-1085. doi:10.1002/gps

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179-195). Oxford, NY: Oxford University Press.

Mitchell, K. J., Johnson, M. K., & Mather, M. (2003). Source monitoring and suggestibility to misinformation: Adult age-related differences. *Applied Cognitive Psychology, 17*, 107-119. doi:10.1002/acp.857

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100. doi:10.1006/cogp.1999.0734

Mueller-Johnson, K., & Ceci, S. J. (2004). Memory and suggestibility in older adults: Live event participation and repeated interview. *Applied Cognitive Psychology, 18*, 1109-1127. doi:10.1002/acp1078

Multhaup, K. S. (1995). Aging, source, and decision criteria: When false fame errors do and do not occur. *Psychology and Aging, 10*, 492-497. doi:10.1037/0882-7974.10.3.492

Multhaup, K. S., & Conner, C. A. (2002). The effects of considering nonlist sources on the Deese-Roediger-McDermott memory illusion. *Journal of Memory and Language, 47*, 214-228. doi:10.1016/S0749-596X(02)00007-4

Multhaup, K. S., De Leonardis, D. M., & Johnson, M. K. (1999). Source memory and eyewitness suggestibility in older adults. *Journal of General Psychology, 126*, 74-84. doi:10.1080/00221309909595352

Münsterberg, H. (1908). *On the witness stand: Essays on psychology and crime*. Garden City, NY: Doubleday, Page & Company.

Murphy, K. J., West, R., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2007). Word-list-learning performance in younger and older adults: Intra-individual performance variability and false memory. *Aging, Neuropsychology, and Cognition*, 70-94. doi:10.1080/138255890969726

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*, 695-699. doi:10.1111/j.1532-5415.2005.53221.x

Naveh-Benjamin, M. (1990). Coding of temporal order information: An automatic process? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 117-126. doi:10.1037/0278-7393.16.1.117

Neath, I., & Surprenant, A. M. (2003). *Human memory* (2nd ed.). Belmont, CA: Wadsworth/Thomson.

Neisser, U. (1967). *Cognitive psychology*. New York, NY: Meredith.

Neuschatz, J. S., Benoit, G. E., & Payne, D. G. (2003). Effective warnings in the Deese-Roediger-McDermott false-memory paradigm: The role of identifiability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 35-40. doi:10.1037/0278-7393.29.1.35

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*, 838-848. doi:10.3758/BF03211328

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97*, 316-333. doi:10.1037/0033-2909.97.2.316

Okun, M. A. (1976). Adult age and cautiousness in decision. *Human Development, 19*. doi:10.1159/000271530

Pallant, J. (2007). *SPSS survival manual* (3rd ed.). New South Wales, Australia: Allen & Unwin.

Park, D. C. (2000). The basic mechanisms accounting for age-related decline in cognitive function. In D. Park & N. Schwarz (Eds.), *Cognitive aging: A primer* (pp. 3-22). Philadelphia, PA: Psychology Press, Taylor & Francis.

Park, D. C., & Minear, M. (2004). Cognitive aging: New directions for old theories. In R. A. Dixon, L. Bäckman, & L. Nilsson (Eds.), *New frontiers in cognitive aging* (pp. 19-114). Oxford, NY: Oxford University Press.

Park, D. C., Puglisi, J. T., & Lutz, R. (1982). Spatial memory in older adults: Effects of intentionality. *Journal of Gerontology, 37*, 330-335. doi:10.1093/geronj/37.3.330

Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J. L., Frieske, D., Zwahr, M., et al. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging, 11*, 621-637. doi:10.1037/0882-7974.11.4.621

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35*, 261-285. doi:10.1006/jmla.1996.0015

Persad, C. C., Abeles, N., Zacks, R. T., & Denburg, N. L. (2002). Inhibitory changes after age 60 and their relationship to measures of attention and memory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57*, 223-232. doi:10.1093/geronb/57.3.P223

Peters, M. J. V., Jelicic, M., Haas, N., & Merckelbach, H. (2006). Mild executive dysfunctions in undergraduates are related to recollecting words never presented. *International Journal of Neuroscience, 116*, 1065-1077. doi:10.1080/00207450600808768

Peters, M. J. V., Jelicic, M., Verbeek, H., & Merckelbach, H. (2007). Poor working memory predicts false memories. *The European Journal of Cognitive Psychology, 19*, 213-232. doi:10.1080/09541440600760396

Pezdek, K., & Lam, S. (2007). What research paradigms have cognitive psychologists used to study "false memory," and what are the implications of these choices? *Consciousness and Cognition, 16*, 2-17. doi:10.1016/j.concog.2005.06.006

Pfeiffer, E. (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society, 23*, 433-441.

Pierce, B. H., Sullivan, A. L., Schacter, D. L., & Budson, A. E. (2005). Comparing source-based and gist-based false recognition in aging and Alzheimer's disease. *Neuropsychology, 19*, 411-419. doi:10.1037/0894-4105.19.4.411

Pinneo, L. R. (1966). On noise in the nervous system. *Psychological Review, 73*, 242-247. doi:10.1037/h0023240

Poon, L. W., & Fozard, J. L. (1980). Age and word frequency effects in continuous recognition memory. *Journal of Gerontology, 35*, 77-86. doi:10.1093/geronj/35.1.77

Rabinowitz, J. C., Craik, F. I. M., & Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology, 36*, 325-344. doi:10.1037/h0080643

Read, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin and Review, 3*, 105-111. doi:10.3758/BF03210749

Reyna, V. F. (2000). Fuzzy-trace theory and source monitoring: An evaluation of theory and false-memory data. *Learning and Individual Differences, 12*, 163-175. doi:10.1016/S1041-6080(01)00034-6

Reyna, V. F., & Brainerd, C. J. (1995a). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1-75. doi:10.1016/1041-6080(95)90031-4

Reyna, V. F., & Brainerd, C. J. (1995b). Fuzzy-trace theory: Some foundational issues. *Learning and Individual Differences, 7*, 145-162. doi:10.1016/1041-6080(95)90028-4

Reyna, V. F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology, 30*, 178-191. doi:10.1037/0012-1649.30.2.178

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9*, 95-123. doi:10.1016/S1041-6080(97)90002-9

Reyna, V. F., & Titcomb, A. L. (1997). Constraints on the suggestibility of eyewitness testimony: A fuzzy-trace theory analysis. In D. G. Payne & F. G. Conrad (Eds.), *Intersections in basic and applied memory research* (pp. 157-174). Hillsdale, NJ: Erlbaum.

Robinson, K. J. (1998). *The automatic mechanisms of false memory* (Unpublished master's thesis). Washington University, St. Louis, MO.

Robinson, K. J., & Roediger, H. L., III. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*, 231-237. doi:10.1111/j.1467-9280.1997.tb00417.x

Roediger, H. L., III, Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger, III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder*. Washington, DC: American Psychological Association.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803-814. doi:10.1037/0278-7393.21.4.803

Roediger, H. L., III, & McDermott, K. B. (1996). False perceptions of false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 814-816. doi:10.1037/0278-7393.22.3.814

Roediger, H. L., III, & McDermott, K. B. (1999). False alarms about false memories. *Psychological Review, 106*, 406-410. doi:10.1037/0033-295X.106.2.406

Roediger, H. L., III, & McDermott, K. B. (2000a). Distortions of memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 149-162). Oxford, NY: Oxford University Press.

Roediger, H. L., III, & McDermott, K. B. (2000b). Tricks of memory. *Current Directions in Psychological Science, 9*, 123-127. doi:10.1111/1467-8721.00075

Roediger, H. L., III, McDermott, K. B., Pisoni, D. B., & Gallo, D. A. (2004). Illusory recollection of voices. *Memory, 12*, 586-602. doi:10.1080/09658210344000125

Roediger, H. L., III, Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin and Review, 8*, 385-407. doi:10.3758/BF03196177

Rosa, N. M., & Gutchess, A. H. (2011). Source memory for action in young and older adults: Self vs. Close or unknown others. *Psychology and Aging, 26*, 625-630. doi:10.1037/a0022827

Rossetti, H. C., Lacritz, L. H., Cullum, C. M., & Weiner, M. F. (2011). Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology, 77*, 1272-1275. doi:10.1212/WNL.0b013e318230208a

Rotello, C. M., & Heit, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory and Language, 40*, 432-453. doi:10.1006/jmla.1998.2623

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28*, 907-922. doi:10.3758/BF03209339

Rowe, G., Valderrama, A., Hasher, L., & Lenartowicz, A. (2006). Attentional disregulation: A benefit for implicit memory. *Psychology and Aging, 21*, 826-830. doi:10.1037/0882-7974.21.4.826

Rybash, J. M., & Hrubi-Bopp, K. L. (2000). Source monitoring and false recollection: A life span developmental perspective. *Experimental Aging Research, 26*, 75-87. doi:10.1080/036107300243696

Salthouse, T. A. (1990). Working memory as a processing resource in cognitive aging. *Developmental Review, 10*, 101-124. doi:10.1016/0273-2297(90)90006-P

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*, 403-428. doi:10.1037/0033-295X.103.3.403

Salthouse, T. A. (2010). *Major issues in cognitive aging*. New York, NY: Oxford University Press.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*, 763-776. doi:10.1037/0012-1649.27.5.763

Salthouse, T. A., Kausler, D. H., & Saults, J. S. (1988). Utilization of path-analytic procedures to investigate the role of processing resources in cognitive aging. *Psychology and Aging, 3*, 158-166. doi:10.1037/0882-7974.3.2.158

Schacter, D. L. (1995). Memory distortion: History and current status. In D. L. Schacter & J. T. Coyle (Eds.), *Memory distortion: How minds, brains, and societies reconstruct the past* (pp. 1-46). London, England: Harvard University Press.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1-24. doi:10.1006/jmla.1998.2611

Schacter, D. L., Kaszniak, A. W., Kihlstrom, J. F., & Valdiserri, M. (1991). The relation between source memory and aging. *Psychology and Aging, 6*, 559-568. doi:10.1037/0882-7974.6.4.559

Schacter, D. L., Koutstaal, W., & Norman, K. A. (1997). False memories and aging. *Trends in Cognitive Sciences, 1*, 229-236. doi:10.1016/S1364-6613(97)01068-1

Schacter, D. L., Osowiecki, D., Kaszniak, A. W., Kihlstrom, J. F., & Valdiserri, M. (1994). Source memory: Extending the boundaries of age-related deficits. *Psychology and Aging, 9*, 81-89. doi:10.1037/0882-7974.9.1.81

Schacter, D. L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language, 35*, 319-334. doi:10.1006/jmla.1996.0018

Schelstraete, M.-A., & Hupet, M. (2002). Cognitive aging and inhibitory efficiency in the Daneman and Carpenter's working memory task. *Experimental Aging Research*, 269-279. doi:10.1080/03610730290080326

Schooler, J. W., Bendiksen, M., & Ambadar, Z. (1997). Taking the middle line: Can we accommodate both fabricated and recovered memories of sexual abuse? In M. A. Conway (Ed.), *Recovered memories and false memories* (pp. 251-292). Oxford, NY: Oxford University Press.

Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science, 9*, 20-26. doi:10.1111/1467-9280.00004

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology, 46*, 561-584. doi:10.1146/annurev.ps.46.020195.003021

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127-190. doi:10.1037/0033-295X.84.2.127

Sim, R. (2010a). [False recognition and the illusory truth effect across age groups]. Unpublished raw data.

Sim, R. (2010b). *Memory mistakes and ageing: How susceptibility to false recognition and the illusory truth effect changes across the lifespan* (Master's thesis, Massey University, Palmerston North, New Zealand). Retrieved from http://hdl.handle.net/10179/2199

Skinner, E. I., & Fernandes, M. A. (2009). Illusory recollection in older adults and younger adults under divided attention. *Psychology and Aging, 24*, 211-216. doi:10.1037/a0014177

Smith, T., Gildeh, N., & Holmes, C. (2007). The Montreal Cognitive Assessment: Validity and utility in a memory clinic setting. *The Canadian Journal of Psychiatry, 52*, 329-332. Retrieved from http://www.stvincents.ie/dynamic/file/montreal_reference.pdf

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34. doi:10.1037/0096-3445.117.1.34

Sorabji, R. (1972). *Aristotle on memory*. London, England: Duckworth.

Sothieson, V. (2010). *Comparing two brief cognitive screening measures in a New Zealand sample* (Unpublished honour's dissertation). Massey University, Auckland, New Zealand.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging, 10*, 527-539. doi:10.1037/0882-7974.10.4.527

Spicer, J. (2005). *Making sense of multivariate data analysis*. Thousand Oaks, CA: Sage.

Stadler, M. A., Roediger, H. L., III, & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*, 494-500. doi:10.3758/BF03211543

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*, 137-149. doi:10.3758/BF03207704

Statistics New Zealand. (2000). *Population ageing in New Zealand*. Retrieved from http://www.stats.govt.nz/browse_for_stats/people_and_communities/older_people/pop-ageing-in-nz.aspx.

Stevens, J. P. (1980). Power of the multivariate analysis of variance tests. *Psychological Bulletin, 88*, 728-737. doi:10.1037/0033-2909.88.3.728

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York, NY: Routledge.

Strauss, H.-M., Leathem, J., Humpries, S., & Podd, J. (2012). The use of brief screening instruments for age-related cognitive impairment in New Zealand. *New Zealand Journal of Psychology, 41*, 13-22. Retrieved from http://content.ebscohost.com.ezproxy.massey.ac.nz/ContentServer.asp?T=P&P=AN&K=89334647&S=R&D=pbh&EbscoContent=dGJyMNXb4kSeqLc4zdnyOLCmr0yeqLBSr624TLWWxWXS&ContentCustomer=dGJyMPGprkq1qa5OuePfgeyx43zx

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1379-1396. doi:10.1037/0278-7393.24.6.1379

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General, 18*, 643-622. doi:10.1037/h0054651

Suengas, A. G., Gallego-Largo, T. R., & Simon, T. (2010). Age-related changes in recognition and response criterion. *Spanish Journal of Psychology, 13*, 557-571. doi:10.1017/S1138741600002249

Sugrue, K., & Hayne, H. (2006). False memories produced by children and adults in the DRM paradigm. *Applied Cognitive Psychology, 20*, 625-631. doi:10.1002/acp.1214

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson Education.

Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language, 52*, 71-86. doi:10.1016/j.jml.2004.08.001

Titcomb, A. L., & Reyna, V. F. (1995). Memory interference and misinformation effects. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 263-294). San Diego, CA: Academic Press.

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

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13*, 230-241. doi:10.1037/0882-7974.13.2.230

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70*, 122-129. doi:10.1037/h0022014

Unsworth, N., & Brewer, G. A. (2010). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language, 62*, 19-34. doi:10.1016/j.jml.2009.08.002

Vasey, M. W., & Thayer, J. F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution. *Psychophysiology, 24*, 479-486. doi:10.1111/j.1469-8986.1987.tb00324.x

Verhaeghen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science, 20*, 174-180. doi:10.1177/0963721411408772

Verhaeghen, P., Marcoen, A., & Goossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. *Journal of Gerontology, 48*, P157-P171. doi:10.1093/geronj/48.4.P157

Wade, K. A., Green, S. L., & Nash, R. A. (2010). Can fabricated evidence induce false eyewitness testimony? *Applied Cognitive Psychology, 24*, 899-908. doi:10.1002/acp.1607

Wade, K. A., Sharman, S. J., Garry, M., Memon, A., Mazzoni, G., Merckelbach, H., et al. (2007). False claims about false memory research. *Consciousness and Cognition, 16*, 18-28. doi:10.1016/j.concog.2006.07.001

Waldie, B. D., & See, S. T. K. (2003). Remembering words never presented: False memory effects in dementia of the Alzheimer type. *Aging, Neuropsychology, and Cognition, 10*, 281-297. doi:10.1076/anec.10.4.281.28969

Waldron-Perrine, B., & Axelrod, B. N. (2012). Determining an appropriate cutting score for indication of impairment on the Montreal Cognitive Assessment. *International Journal of Geriatric Psychiatry, 27*, 1189-1194. doi:10.1002/gps.3768

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 1*, 442-452. doi:10.1037/0278-7393.1.4.442

Watson, J. M., Balota, D. A., & Sergent-Marshall, S. D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with dementia of the Alzheimer's type. *Neuropsychology, 15*, 254-267. doi:10.1037/0894-4105.15.2.254

Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. A. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 76-85. doi:10.1037/0278-7393.31.1.76

Watson, J. M., McDermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger-McDermott paradigm: Assessing the combined influence of practice and warnings in young and old adults. *Memory & Cognition, 32*, 135-141. doi:10.3758/BF03195826

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36*, 1546-1557. doi:10.1037/0022-3514.36.12.1546

West, R., & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology, 37*, 179-189. doi:10.1111/1469-8986.3720179

Westerberg, C. E., & Marsolek, C. J. (2003). Sensitivity reductions in false recognition: A measure of false memories with stronger theoretical implications. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 747-759. doi:10.1037/0278-7393.29.5.747

Wickens, T. D., & Hirshman, E. (2000). False memories and statistical decision theory: Comment on Miller and Wolford (1999) and Roediger and McDermott (1999). *Psychological Review, 107*, 377-383. doi:10.1037/0033-295X.107.2.377

Wise, R. A., Pawlenko, N. B., Safer, M. A., & Meyer, D. (2009). What US prosecutors and defence attorneys know and believe about eyewitness testimony. *Applied Cognitive Psychology, 23*, 1266-1281. doi:10.1002/acp.1530

Wixted, J. T., & Stretch, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review, 107*, 368-376. doi:10.1037/0033-295X.107.2.368

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 934-934. doi:10.1037/0278-7393.20.4.934

List of Appendices

List of Tables

List of Figures

Appendix A: Ethics Approval Letter

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA

12 December 2011

Rachael Sim
[redacted]

Dear Rachael

Re: HEC: Southern B Application – 11/73
Do common memory mistakes change with age?

Thank you for your letter dated 2 December 2011.

On behalf of the Massey University Human Ethics Committee: Southern B I am pleased to advise you that the ethics of your application are now approved. Approval is for three years. If this project has not been completed within three years from the date of this letter, reapproval must be requested.

If the nature, content, location, procedures or personnel of your approved application change, please advise the Secretary of the Committee.

Yours sincerely

Dr Nathan Matthews, Acting Chair
Massey University Human Ethics Committee: Southern B

cc Dr Stephen Hill
School of Psychology
PN320

A/Prof John Podd
School of Psychology
PN320

Dr Joanne Taylor
School of Psychology
PN320

A/Prof Mandy Morgan, HoS
School of Psychology
PN320

Massey University Human Ethics Committee
Accredited by the Health Research Council
Research Ethics Office, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand
T +64 6 350 5573 +64 6 350 5575 F +64 6 350 5622
E humanethics@massey.ac.nz animalethics@massey.ac.nz gtc@massey.ac.nz
www.massey.ac.nz

Appendix B: Standard Recognition Test

Instructions:

1. Please indicate (by circling or putting a line through a response) if you:

   Heard the word **(H)**

   Or

   Did not hear the word **(N)**

2. Please indicate (by circling or putting a line through a response) how certain you are that you heard the word played in the list or that you did not hear the word played in the list.

   1 = Very certain **heard**

   2 = Mostly certain **heard**

   3 = Fairly certain **heard**

   4 = Fairly certain **not heard**

   5 = Mostly certain **not heard**

   6 = Very certain **not heard**

Word	Response		Certainty					
	Heard / Not Heard		Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Cage	H	N	1	2	3	4	5	6
Zoo	H	N	1	2	3	4	5	6
Kiwi	H	N	1	2	3	4	5	6
Sign	H	N	1	2	3	4	5	6
Happy	H	N	1	2	3	4	5	6
Lethargic	H	N	1	2	3	4	5	6
Ready	H	N	1	2	3	4	5	6
Capital	H	N	1	2	3	4	5	6
Vegetable	H	N	1	2	3	4	5	6
Charred	H	N	1	2	3	4	5	6
Rough	H	N	1	2	3	4	5	6
Summit	H	N	1	2	3	4	5	6
Yawn	H	N	1	2	3	4	5	6

	Heard / Not Heard		Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Sleep	H	N	1	2	3	4	5	6
Delay	H	N	1	2	3	4	5	6
Red	H	N	1	2	3	4	5	6
Fly	H	N	1	2	3	4	5	6
Sand	H	N	1	2	3	4	5	6
Glacier	H	N	1	2	3	4	5	6
High	H	N	1	2	3	4	5	6
Pole	H	N	1	2	3	4	5	6
Go	H	N	1	2	3	4	5	6
Coaster	H	N	1	2	3	4	5	6
Thief	H	N	1	2	3	4	5	6
Hunt	H	N	1	2	3	4	5	6
Uneven	H	N	1	2	3	4	5	6
Tough	H	N	1	2	3	4	5	6

	Heard / Not Heard		Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Crowded	H	N	1	2	3	4	5	6
Medicine	H	N	1	2	3	4	5	6
Pillow	H	N	1	2	3	4	5	6
Slow	H	N	1	2	3	4	5	6
Boat	H	N	1	2	3	4	5	6
Chicago	H	N	1	2	3	4	5	6
Keys	H	N	1	2	3	4	5	6
Kick	H	N	1	2	3	4	5	6
Mouse	H	N	1	2	3	4	5	6
Building	H	N	1	2	3	4	5	6
Lion	H	N	1	2	3	4	5	6
Mountain	H	N	1	2	3	4	5	6
Dark	H	N	1	2	3	4	5	6
Pear	H	N	1	2	3	4	5	6

	Heard / Not Heard		Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Village	H	N	1	2	3	4	5	6
Tamer	H	N	1	2	3	4	5	6
Turtle	H	N	1	2	3	4	5	6
Black	H	N	1	2	3	4	5	6
Food	H	N	1	2	3	4	5	6
Peak	H	N	1	2	3	4	5	6
Comfort	H	N	1	2	3	4	5	6
Infantry	H	N	1	2	3	4	5	6
Rest	H	N	1	2	3	4	5	6
Juice	H	N	1	2	3	4	5	6
Tired	H	N	1	2	3	4	5	6
Sheep	H	N	1	2	3	4	5	6
Hands	H	N	1	2	3	4	5	6
Soft	H	N	1	2	3	4	5	6

	Heard / Not Heard		Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Ground	H	N	1	2	3	4	5	6
Junk	H	N	1	2	3	4	5	6
Square	H	N	1	2	3	4	5	6
Quick	H	N	1	2	3	4	5	6
Dog	H	N	1	2	3	4	5	6
Muscle	H	N	1	2	3	4	5	6
Needle	H	N	1	2	3	4	5	6
Country	H	N	1	2	3	4	5	6
Bottom	H	N	1	2	3	4	5	6
Fierce	H	N	1	2	3	4	5	6
Nap	H	N	1	2	3	4	5	6
Snooze	H	N	1	2	3	4	5	6
Steep	H	N	1	2	3	4	5	6
Music	H	N	1	2	3	4	5	6

	Heard / Not Heard		Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Bowl	H	N	1	2	3	4	5	6
Doze	H	N	1	2	3	4	5	6
Wait	H	N	1	2	3	4	5	6
Valley	H	N	1	2	3	4	5	6
Lamb	H	N	1	2	3	4	5	6
Cake	H	N	1	2	3	4	5	6
Climber	H	N	1	2	3	4	5	6
Sound	H	N	1	2	3	4	5	6
People	H	N	1	2	3	4	5	6
Fruit	H	N	1	2	3	4	5	6
Sister	H	N	1	2	3	4	5	6
Bumpy	H	N	1	2	3	4	5	6
Snow	H	N	1	2	3	4	5	6
Loaf	H	N	1	2	3	4	5	6

	Heard / Not Heard		Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Brown	H	N	1	2	3	4	5	6
Listless	H	N	1	2	3	4	5	6
Lazy	H	N	1	2	3	4	5	6
Stream	H	N	1	2	3	4	5	6
Blue	H	N	1	2	3	4	5	6
City	H	N	1	2	3	4	5	6
Colour	H	N	1	2	3	4	5	6
Crayon	H	N	1	2	3	4	5	6
Africa	H	N	1	2	3	4	5	6
County	H	N	1	2	3	4	5	6
Wood	H	N	1	2	3	4	5	6
Berry	H	N	1	2	3	4	5	6
Circus	H	N	1	2	3	4	5	6

Appendix C: Source-Monitoring Recognition Test

Instructions:

1. Please indicate (by circling or putting a line through a response) if you:

    Heard the word **(H)**

    Heard the word and thought of it on your own **(H & T)**

    Did not hear the word but thought of it on your own **(Didn't H, but T)**

    Or

    Did not hear or think of the word **(Didn't H or T)**

2. Please indicate (by circling or putting a line through a response) how certain you are that you heard the word played in the list or that you did not hear the word played in the list.

    1 = Very certain **heard**

    2 = Mostly certain **heard**

    3 = Fairly certain **heard**

    4 = Fairly certain **not heard**

    5 = Mostly certain **not heard**

    6 = Very certain **not heard**

Word	Response				Certainty					
	I heard this word	I heard this word and thought of it on my own	I did not hear this word but thought of it on my own	I did not hear or think of this word	Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Cage	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Zoo	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Kiwi	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Sign	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Happy	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Lethargic	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Ready	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Capital	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Vegetable	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Charred	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Rough	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Summit	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6
Yawn	H	H & T	Didn't H, but T	Didn't H or T	1	2	3	4	5	6

	I heard this word	I heard this word and thought of it on my own	I did not hear this word but thought of it on my own	I did not hear or think of this word	Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Sleep	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Delay	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Red	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Fly	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Sand	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Glacier	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
High	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Pole	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Go	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Coaster	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Thief	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Hunt	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Uneven	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Tough	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6

	I heard this word	I heard this word and thought of it on my own	I did not hear this word but thought of it on my own	I did not hear or think of this word	Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Crowded	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Medicine	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Pillow	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Slow	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Boat	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Chicago	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Keys	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Kick	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Mouse	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Building	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Lion	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Mountain	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Dark	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Pear	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6

	I heard this word	I heard this word and thought of it on my own	I did not hear this word but thought of it on my own	I did not hear or think of this word	Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Village	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Tamer	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Turtle	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Black	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Food	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Peak	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Comfort	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Infantry	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Rest	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Juice	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Tired	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Sheep	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Hands	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Soft	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6

	I heard this word	I heard this word and thought of it on my own	I did not hear this word but thought of it on my own	I did not hear or think of this word	Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Ground	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Junk	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Square	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Quick	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Dog	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Muscle	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Needle	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Country	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Bottom	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Fierce	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Nap	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Snooze	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Steep	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Music	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6

	I heard this word	I heard this word and thought of it on my own	I did not hear this word but thought of it on my own	I did not hear or think of this word	Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Bowl	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Doze	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Wait	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Valley	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Lamb	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Cake	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Climber	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Sound	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
People	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Fruit	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Sister	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Bumpy	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Snow	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Loaf	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6

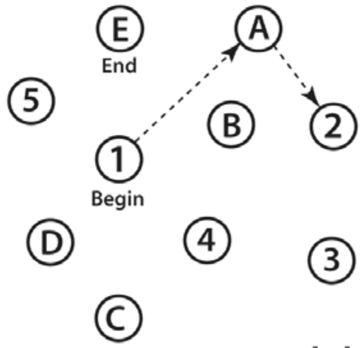
	I heard this word	I heard this word and thought of it on my own	I did not hear this word but thought of it on my own	I did not hear or think of this word	Very certain **heard**	Mostly certain **heard**	Fairly certain **heard**	Fairly certain **not heard**	Mostly certain **not heard**	Very certain **not heard**
Brown	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Listless	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Lazy	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Stream	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Blue	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
City	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Colour	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Crayon	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Africa	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
County	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Wood	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Berry	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6
Circus	H	H & T	Didn’t H, but T	Didn’t H or T	1	2	3	4	5	6

Appendix D: Montreal Cognitive Assessment

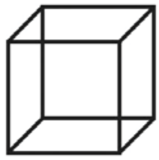
MONTREAL COGNITIVE ASSESSMENT (MOCA)
Version 7.1 Original Version

NAME :
Education :
Sex :
Date of birth :
DATE :

VISUOSPATIAL / EXECUTIVE — POINTS



[ ]



Copy cube

[ ]

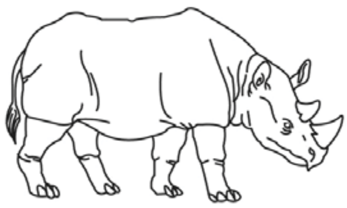
Draw CLOCK (Ten past eleven)
(3 points)

[ ] Contour [ ] Numbers [ ] Hands

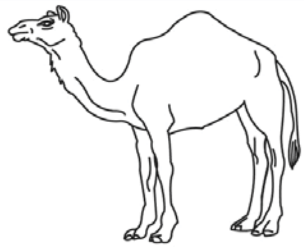
___/5

NAMING

[ ]



[ ]



[ ]

___/3

MEMORY Read list of words, subject must repeat them. Do 2 trials, even if 1st trial is successful. Do a recall after 5 minutes.

	FACE	VELVET	CHURCH	DAISY	RED
1st trial					
2nd trial					

No points

ATTENTION Read list of digits (1 digit/ sec.).

Subject has to repeat them in the forward order [ ] 2 1 8 5 4
Subject has to repeat them in the backward order [ ] 7 4 2

___/2

Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors
[ ] F B A C M N A A J K L B A F A K D E A A A J A M O F A A B

___/1

Serial 7 subtraction starting at 100 [ ] 93 [ ] 86 [ ] 79 [ ] 72 [ ] 65
4 or 5 correct subtractions: **3 pts**, 2 or 3 correct: **2 pts**, 1 correct: **1 pt**, 0 correct: **0 pt**

___/3

LANGUAGE Repeat : I only know that John is the one to help today. [ ]
The cat always hid under the couch when dogs were in the room. [ ]

___/2

Fluency / Name maximum number of words in one minute that begin with the letter F [ ] _____ (N ≥ 11 words)

___/1

ABSTRACTION Similarity between e.g. banana - orange = fruit [ ] train – bicycle [ ] watch - ruler

___/2

DELAYED RECALL	Has to recall words **WITH NO CUE**	FACE [ ]	VELVET [ ]	CHURCH [ ]	DAISY [ ]	RED [ ]	Points for UNCUED recall only
Optional	Category cue						
	Multiple choice cue						

___/5

ORIENTATION [ ] Date [ ] Month [ ] Year [ ] Day [ ] Place [ ] City

___/6

 www.mocatest.org Normal ≥26 / 30

TOTAL ___/30
Add 1 point if ≤ 12 yr edu

Administered by: ____________________

MoCA Administration Instructions

1. Alternating Trail Making

Administration: The examiner instructs the subject: "Please draw a line, going from a number to a letter in ascending order. Begin here [point to (1)] and draw a line from 1 then to A then to 2 and so on. End here [point to (E)]."

Scoring: Allocate one point if the subject successfully draws the following pattern: 1 –A- 2- B- 3- C- 4- D- 5- E, without drawing any lines that cross. Any error that is not immediately self-corrected earns a score of 0.

2. Visuoconstructional Skills (Cube)

Administration: The examiner gives the following instructions, pointing to the cube: “Copy this drawing as accurately as you can, in the space below”.

Scoring: One point is allocated for a correctly executed drawing.

- Drawing must be three-dimensional
- All lines are drawn
- No line is added
- The horizontal lines are relatively parallel
- The object must be clearly rectangular (i.e., the shorter vertical sides cannot be more than ¾ of the length of the longer horizontal lines).

A point is not assigned if any of the above-criteria are not met.

3. Visuoconstructional Skills (Clock)

Administration: Indicate the right third of the space and give the following instructions: “Draw a clock. Put in all the numbers and set the time to 10 past 11”.

Scoring: One point is allocated for each of the following three criteria:

- Contour (1 point): the clock face must be a circle with only minor distortion acceptable (e.g., slight imperfection on closing the circle);
- Numbers (1 point): all clock numbers must be present with no additional numbers; numbers must be in the correct order and placed in the approximate quadrants on the clock face; Roman numerals are acceptable; numbers can be placed outside the circle contour;

- Hands (1 point): there must be two hands jointly indicating the correct time; the hour hand must be clearly shorter than the minute hand; hands must be centred within the clock face with their junction close to the clock centre.

A point is not assigned for a given element if any of the above-criteria are not met.

4. Naming

Administration: Beginning on the left, point to each figure and say: “Tell me the name of this animal”.

Scoring: One point each is given for the following responses:

I. Lion
II. Rhinoceros or Rhino
III. Camel or Dromedary

5. Memory

Administration: The examiner reads a list of 5 words at a rate of one per second, giving the following instructions: “This is a memory test. I am going to read a list of words that you will have to remember now and later on. Listen carefully. When I am through, tell me as many words as you can remember. It doesn’t matter in what order you say them”. Mark a check in the allocated space for each word the subject produces on this first trial. When the subject indicates that (s)he has finished (has recalled all words), or can recall no more words, read the list a second time with the following instructions: “I am going to read the same list for a second time. Try to remember and tell me as many words as you can, including words you said the first time.” Put a check in the allocated space for each word the subject recalls after the second trial.

At the end of the second trial, inform the subject that (s)he will be asked to recall these words again by saying, “I will ask you to recall those words again at the end of the test.”

Scoring: No points are given for Trials One and Two.

6. Attention

Forward Digit Span

Administration: Give the following instruction: "I am going to say some numbers and when I am through, repeat them to me exactly as I said them". Read the five number sequence at a rate of one digit per second.

Scoring: Allocate one point for each sequence correctly repeated, (N.B.: the correct response for the backwards trial is 2-4-7).

Backward Digit Span

Administration: Give the following instruction: "Now I am going to say some more numbers, but when I am through you must repeat them to me in the backwards order." Read the three number sequence at a rate of one digit per second.

Scoring: Allocate one point for each sequence correctly repeated, (N.B.: the correct response for the backwards trial is 2-4-7).

Vigilance

Administration: The examiner reads the list of letters at a rate of one per second, after giving the following instruction: "I am going to read a sequence of letters. Every time I say the letter A, tap your hand once. If I say a different letter, do not tap your hand".

Scoring: Give one point if there is zero to one errors (an error is a tap on a wrong letter or a failure to tap on letter A).

Serial 7s

Administration: The examiner gives the following instruction: "Now, I will ask you to count by subtracting seven from 100, and then, keep subtracting seven from your answer until I tell you to stop." Give this instruction twice if necessary.

Scoring: This item is scored out of 3 points. Give no (0) points for no correct subtractions, 1 point for one correction subtraction, 2 points for two-to-three correct subtractions, and 3 points if the participant successfully makes four or five correct subtractions. Count each correct subtraction of 7 beginning at 100. Each subtraction is evaluated independently; that is, if the participant responds with an incorrect number but continues to correctly subtract 7 from it, give a point for each correct subtraction. For example, a participant may respond "92 – 85 – 78 – 71 – 64" where the "92" is incorrect, but all subsequent numbers are subtracted correctly. This is one error and the item would be given a score of 3.

7. Sentence Repetition

Administration: The examiner gives the following instructions: “I am going to read you a sentence. Repeat it after me, exactly as I say it [pause]: I only know that John is the one to help today.” Following the response, say: “Now I am going to read you another sentence. Repeat it after me, exactly as I say it [pause]: The cat always hid under the couch when dogs were in the room.”

Scoring: Allocate 1 point for each sentence correctly repeated. Repetition must be exact. Be alert for errors that are omissions (e.g., omitting "only", "always") and substitutions/additions (e.g., "John is the one who helped today;" substituting "hides" for "hid", altering plurals, etc.).

8. Verbal Fluency

Administration: The examiner gives the following instruction: “Tell me as many words as you can think of that begin with a certain letter of the alphabet that I will tell you in a moment. You can say any kind of word you want, except for proper nouns (like Bob or Boston), numbers, or words that begin with the same sound but have a different suffix, for example, love, lover, loving. I will tell you to stop after one minute. Are you ready? [Pause] Now, tell me as many words as you can think of that begin with the letter F. [time for 60 sec]. Stop.”

Scoring: Allocate one point if the subject generates 11 words or more in 60 sec. Record the subject’s response in the bottom or side margins.

9. Abstraction

Administration: The examiner asks the subject to explain what each pair of words has in common, starting with the example: “Tell me how an orange and a banana are alike”. If the subject answers in a concrete manner, then say only one additional time: “Tell me another way in which those items are alike”. If the subject does not give the appropriate response (fruit), say, “Yes, and they are also both fruit.” Do not give any additional instructions or clarification. After the practice trial, say: “Now, tell me how a train and a bicycle are alike”. Following the response, administer the second trial, saying: “Now tell me how a ruler and a watch are alike”. Do not give any additional instructions or prompts.

Scoring: Only the last two item pairs are scored. Give 1 point to each item pair correctly answered. The following responses are acceptable:

- Train-bicycle = means of transportation, means of travelling, you take trips in both;
- Ruler-watch = measuring instruments, used to measure.

The following responses are not acceptable: Train-bicycle = they have wheels; Ruler-watch = they have numbers.

10. Delayed Recall

Administration: The examiner gives the following instruction: “I read some words to you earlier, which I asked you to remember. Tell me as many of those words as you can remember.” Make a check mark (√) for each of the words correctly recalled spontaneously without any cues, in the allocated space.

Scoring: Allocate 1 point for each word recalled freely without any cues.

Optional: Following the delayed free recall trial, prompt the subject with the semantic category cue provided below for any word not recalled. Make a check mark (√) in the allocated space if the subject remembered the word with the help of a category or multiple-choice cue. Prompt all non-recalled words in this manner. If the subject does not recall the word after the category cue, give him/her a multiple choice trial, using the following example instruction, “Which of the following words do you think it was, NOSE, FACE, or HAND?” Use the following category and/or multiple-choice cues for each word, when appropriate:

Item	Category Cue	Multiple choice
FACE	Part of the body	Nose, face, hand
VELVET	Type of fabric	Denim, cotton, velvet
CHURCH	Type of building	Church, school, hospital
DAISY	Type of flower	Rose, daisy, tulip
RED	A colour	Red, blue, green

Scoring: No points are allocated for words recalled with a cue. A cue is used for clinical information purposes only and can give the test interpreter additional information about the type of memory disorder. For memory deficits due to retrieval failures, performance can be improved with a cue. For memory deficits due to encoding failures, performance does not improve with a cue.

11. Orientation

Administration: The examiner gives the following instructions: “Tell me the date today”. If the subject does not give a complete answer, then prompt accordingly by saying: “Tell me the [year, month, exact date, and day of the week].” Then say: “Now, tell me the name of this place, and which city it is in.”

Scoring: Give one point for each item correctly answered. The subject must tell the exact date and the exact place (name of hospital, clinic, office). No points are allocated if subject makes an error of one day for the day and date.

TOTAL SCORE: Sum all subscores listed on the right-hand side. Add one point for an individual who has 12 years or fewer of formal education, for a possible maximum of 30 points. A final total score of 26 and above is considered normal.

Appendix E: Permission to Use the Montreal Cognitive Assessment

Subject: RE: Permission to use the MoCA
From: Info-MoCA <info@mocatest.org>
Date: 12/10/2011 07:01
To: 'Rachael Sim' <rachael_sim@yahoo.com>

Good afternoon,

You are welcome to use the MoCA in your study as described below with no further permission requirements if it is not industry funded.

Any modification to the MoCA ©/ Instructions, requires prior written approval by copyright owner.

All the best,

Tina Brosseau
Projects & Development Manager
Center for Diagnosis & Research on Alzheimer's disease (CEDRA)
Phone: (450) 672-9637 / Fax: (450) 672-1443
www.cedra.ca / www.mocatest.org

Subject: Re: Permission to use the MoCA
From: Rachael Sim [mailto:rachael_sim@yahoo.com]
Sent: 3/10/2011 14:37
To: info@mocatest.org

Hi Tina,

The title is False Memories and Ageing.
The project aims to examine age-related differences in false memories and if false recognition is best explained by a fuzzy-trace framework, an activation-monitoring framework, or signal detection ability.

Approximately 80 participants will take part in the research, possibly up to 120 if time permits and 80 doesn't allow us to find significant results due to small sample and effect sizes. The MoCA will be administered to each participant once. This project is not industry funded.

Thank you
Rachael Sim

Appendix F: Participant Information Sheet

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Do Common Memory Mistakes Change with Age?

INFORMATION SHEET

Who is doing this project?

Hello, my name is Rachael Sim and I am currently working on my Doctoral degree in Clinical Psychology. This research is part of the requirements to complete this qualification from Massey University. The main supervisor for my research is Dr. Stephen Hill, who is based at Massey University in Palmerston North. If at any stage you have any questions regarding this research, please feel free to contact myself or Stephen Hill in the following ways:

Rachael Sim

Phone 027 6252277
06-3269137
Email rachael_sim@yahoo.com

Dr. Stephen Hill

Phone (06) 356 9099 ext. 7566
Email s.r.hill@massey.ac.nz
Mail School of Psychology
Massey University
Private Bag 11-222
Palmerston north

What is the project about?

I am interested in the way you remember lists of words. We are not only interested in how many you accurately remember; we're also interested in the kinds of mistakes you make, and what words you remember better than others. Memory mistakes are normal in everyone and the mistakes we make provide interesting information about the way normal, everyday memory works.

This is an invitation for you to participate in this project. If you wish to participate please contact me (Rachael) using the contact information above and we can arrange a time and place so you can take part in the project.

What does the study involve?

The project consists of the following three phases:

- You will be asked to complete a brief cognitive assessment measure.
- You will then be asked to listen to a recording of a list of approximately 120 words, which will be immediately followed by a recognition test.
- Lastly, you will be asked to complete a very short questionnaire.

The time involved for the entire process, including, reading information sheets, signing consent forms, and the final debrief, is about 20-40 minutes.

How are participants being identified and recruited?

Approximately 40 participants aged 16-30 and 40 participants aged 75-80 are being asked to voluntarily participate in this research. A total of 80 participants are requir ed because we are comparing the two age groups and we need approximately 40 in each group to detect the patterns and effects in the data.

Participants are being recruited a number of ways, including, through community groups, undergraduate Massey University classes, and by word of mouth. The requirements to participate are that you are aged 16-30 or 75-80, are proficient in the English language, and have reasonable ability to read, write, and hear (including with the use of aids).

Each participant will be reimbursed with a $10 supermarket or petrol voucher for their time.

There should be no discomfort or risks to participants as a result of participation. However, there will be a debriefing period after participation should any problems arise.

What will happen with the information?

The responses given during the three phases will be recorded and collated with those of other participants'. The collated results will then be used to evaluate the study questions discussed earlier in this form. All information given will be kept confidential, and individual results will not be personally identifiable in the research findings. Data will be kept separate from consent forms, and both will be kept securely for a period of five years before being destroyed.

The consent form offers a space where you can provide your personal e-mail or postal address so that we can provide you with a written summary of the research findings.

Your right as a participant

You are under no obligation to accept this invitation. If you decide to participate, you have the right to:

- Decline to answer any particular question.
- Withdraw from the study at any time.
- Ask any questions about the study at any time during participation.
- Provide information on the understanding that your name will not be used unless you give permission to the researcher.
- Be given access to a summary of the project findings when it is concluded.

For student participants, neither grades nor academic relationships with the School of Psychology or members of staff will be affected by your refusal or agreement to participate.

Ethical approval information

This project has been reviewed and approved by the Massey University Human Ethics Committee: Southern B, Application 11/73. If you have any concerns about the conduct of this research, please contact Dr Nathan Matthews, Acting Chair, Massey University Human Ethics Committee: Southern B, telephone 06 350 5799 x 8729, email humanethicsouthb@massey.ac.nz.

Thank you for your interest in this study and taking the time to read this.

Appendix G: Participant Consent Form

Do Common Memory Mistakes Change with Age?

PARTICIPANT CONSENT FORM

I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

I agree to participate in this study under the conditions set out in the Information Sheet.

Signature: .... **Date:** ....

Full Name (printed) ....

I would like to be sent a summary of the overall research findings when they become available.

(Please tick one) ☐ via e-mail

☐ via post

(Please write clearly your e-mail or postal address)

E-mail ____

Address ____

____

____

....

Appendix H: Participant Questionnaire

Do Common Memory Mistakes Change with Age?

Participant Questionnaire

Please clearly write what the time is ______ (hour) ______ (min) AM / PM (circle one)

Please clearly write your age (in years) ______________________

Do you identify as (Please tick one)

Female ☐

Male ☐

Please identify the levels of formal education you hold — tick one

	tick one
Less than NCEA level 1, 5th form certificate, or 11 years of education	☐
NCEA level 1, 5th form certificate, or 11 years of education	☐
NCEA level 2, 6th form certificate, or 12 years of education	☐
NCEA level 3, 7th form (higher school) certificate, bursary, or 13 years of education	☐
Technical institute or apprenticeship	☐
University diploma or degree	☐
Post-graduate diploma or degree	☐

Appendix I: False Memory Task Instructions

Standard Test Instructions

1. In this task you will first hear 120 words that I want you to try and remember. It's not one long list, but 8 shorter lists played one after the other, and each is separated with a bell sound. Therefore, you'll hear about 15 words, then the bell sound, and then the next list will start. Until you've heard a total of 120 words.
2. At the end of the list I will hand you a memory test with 96 words.
3. This is an example of what the test will look like [put the old/new sample in front of them use the sample when instructing participant].
4. This task is likely something you've never come across before, so please ask questions if you don't understand.
5. With each word I first want you to decide if you heard the word played in the list or you did not hear the word played in the list. Either circle or put a line through your answer.
6. Then I want you to decide how certain you are that you heard the word or did not hear the word. For example if you heard the word 'do' in the list you would choose if you were: very certain it was heard, which means you're absolutely certain the word was played in the list; mostly certain it was heard, which means you're not absolutely certain but you're certain the word was played in the list; or fairly certain it was heard, which means you're only somewhat certain the word was played in the list.
7. And then you move to the next word. Just remember if you decide you heard the word then you rate how certain you are the word was heard. Whereas, if you decide you did not hear the word then you rate how certain you are the word was not heard.
8. Note that the numbers here don't mean anything; they're just so it's easier for me to enter the data into a spreadsheet, but do make sure your aware of what certainty your choosing with each word.
9. The test has no time limit, but it shouldn't take any longer than 10-15 minutes.
10. Do you have any questions?
11. Do you understand what you have to do?

12. When I give you the test I'll also place a copy of the basic instructions in front of you so you can refer to them if you need too.
13. Great, when you're ready and I'll play the list, once it's finished I'll give you the test and you can start it immediately.

Source-Monitoring Test Instructions

1. In this task you will first hear 120 words that I want you to try and remember. It's not one long list, but 8 shorter lists played one after the other, and each is separated with a bell sound. Therefore, you'll hear about 15 words, then the bell sound, and then the next list will start. Until you've heard a total of 120 words.
2. At the end of the list I will hand you a memory test with 96 words.
3. This is an example of what the test will look like [put the source sample in front of them use the sample when instructing participant].
4. This task is likely something you've never come across before, so please ask questions if you don't understand.
5. The first thing I'll explain is the response options. There are 2 factors to consider for each word, the first is if you heard the word played in the list or you did not hear the word played in the list. The 2nd is a little more difficult. It's if you thought of the word on your own or not, that means that you remember thinking of the word on your own, or it came to mind, as you were listening to the list and the word may or may not have been played in the list.
6. With each word I first want you to decide if you: heard the word, heard the word and thought of it on your own, did not hear the word but thought of it on our own, or you did not hear the word or think of it on your own. Either circle or put a line through your answer.
7. Then I want you to decide how certain you are that you heard the word or did not hear the word. For example if you heard the word 'do' in the list you would choose if you were: very certain it was heard, which means you're absolutely certain the word was played in the list; mostly certain it was heard, which means you're not absolutely certain but you're certain the word was played in the list; or fairly certain it was heard, which means you're only somewhat certain the word was played in the list.

8. And then you move to the next word. Just remember if you decide you heard the word then you rate how certain you are the word was heard. Whereas, if you decide you did not hear the word then you rate how certain you are the word was not heard.
9. Note that the numbers here don't mean anything; they're just so it's easier for me to enter the data into a spreadsheet, but do make sure your aware of what certainty your choosing with each word.
10. The test has no time limit, but it shouldn't take any longer than 10-15 minutes.
11. Do you have any questions?
12. Do you understand what you have to do?
13. When I give you the test I'll also place a copy of the basic instructions in front of you so you can refer to them if you need too.
14. Great, when you're ready and I'll play the list, once it's finished I'll give you the test and you can start it immediately.

Appendix J: Chapter 5 Normality, Correlation, and Sphericity Test Results

Table J1

Kolmogorov-Smirnov Tests for Recognition Proportions for the Age and Test Condition Groups

	Young				Old			
	Standard task		Source task		Standard task		Source task	
Variable	*D*	*p*	*D*	*p*	*D*	*p*	*D*	*p*
Critical Lure	.27	<.01	.20	.03	.21	.02	.20	.04
Weak Lure	.16	>.20	.17	.16	.08	>.20	.15	>.20
Unrelated	.21	.02	.26	<.01	.33	<.01	.26	<.01
Presented	.16	.17	.15	>.20	.19	.04	.16	.18

Note. For all calculations *df* = 20.

Table J2

Skewness and Kurtosis Z-scores for Recognition Proportions for the Age and Test Condition Groups

	Young				Old			
	Standard task		Source task		Standard task		Source task	
Variable	z_{skewness}	z_{kurtosis}	z_{skewness}	z_{kurtosis}	z_{skewness}	z_{kurtosis}	z_{skewness}	z_{kurtosis}
Critical Lure	-1.26	.45	.48	1.47	-2.23	1.17	-1.02	-.56
Weak Lure	1.23	-.63	1.98	1.76	.88	.20	1.27	-.50
Unrelated	2.48	1.73	3.00	3.27	4.63	5.89	3.48	3.40
Presented	1.16	.85	.59	-1.14	-1.78	.81	-.67	-1.10

Table J3

Kolmogorov-Smirnov Tests for Certainty Ratings for the Age and Test Condition Groups

	Young				Old			
	Standard task		Source task		Standard task		Source task	
Rating	*D*	*p*	*D*	*p*	*D*	*p*	*D*	*p*
Critical Lure	.22	.02	.16	>.20	.20	.03	.20	.03
Weak Lure	.16	>.20	.15	>.20	.11	>.20	.18	.12
Presented	.17	.14	.11	>.20	.20	.03	.17	.11

Note. For the above tests *df* = 20, except for the following: *df* = 18 for younger adults in the source condition critical lure ratings and older adults in the standard condition weak lure ratings, and *df* = 19 for younger adults in the standard condition weak lure ratings.

Table J4

Skewness and Kurtosis Z-scores for Certainty Ratings for the Age and Test Condition Groups

	Young				Old			
	Standard task		Source task		Standard task		Source task	
Variable	z_{skewness}	z_{kurtosis}	z_{skewness}	z_{kurtosis}	z_{skewness}	z_{kurtosis}	z_{skewness}	z_{kurtosis}
Critical Lure	1.52	-.42	2.09	1.73	2.61	2.01	3.17	2.54
Weak Lure	-.19	-1.16	.11	-1.06	.61	-.96	1.15	-.83
Presented	-.59	-.63	-.41	.10	2.75	2.94	2.88	2.77

Table J5

Kolmogorov-Smirnov Tests for the Source-Monitoring Response Options for the Age Groups

	Young		Old	
Variable	*D*	*p*	*D*	*p*
Critical lure response options				
Heard only	.19	.07	.17	.51
Heard and thought	.31	<.01	.48	<.01
Thought only	.17	.12	.29	<.01
Neither heard/thought	.24	<.01	.23	<.01
Weak lure response options				
Heard only	.17	.13	.20	.03
Heard and thought	.25	<.01	.40	<.01
Thought only	.17	.14	.23	<.01
Neither heard/thought	.13	>.20	.11	>.20
Unrelated item response options				
Heard only	.33	<.01	.24	<.01
Heard and thought	.48	<.01	.52	<.01
Thought only	.32	<.01	.33	<.01
Neither heard/thought	.26	<.01	.25	<.01
Presented item response options				
Heard only	.12	>.20	.17	.13
Heard and thought	.25	<.01	.35	<.01
Thought only	.16	.19	.31	<.01
Neither heard/thought	.10	>.20	.21	.02

Note. For all calculations *df* = 20.

Table J6

Skewness and Kurtosis Z-scores for the Source-Monitoring Response Options for the Age Groups

Variable	Young		Old	
	$z_{skewness}$	$z_{kurtosis}$	$z_{skewness}$	$z_{kurtosis}$
Critical lure response options				
Heard only	-.30	-1.54	-.88	-.63
Heard and thought	1.39	-.45	7.38	15.27
Thought only	.99	-.83	2.36	1.15
Neither heard/thought	1.02	-.80	2.57	.82
Weak lure response options				
Heard only	1.63	-.45	2.21	.39
Heard and thought	1.69	-.41	3.90	3.20
Thought only	.67	-1.15	2.43	.35
Neither heard/thought	.09	1.07	-.72	-.53
Unrelated item response options				
Heard only	4.00	4.82	4.29	6.35
Heard and thought	4.37	4.69	6.71	11.98
Thought only	4.39	5.52	7.27	15.30
Neither heard/thought	-2.29	1.23	2.79	1.34
Presented item response options				
Heard only	.91	-.52	-.28	-1.20
Heard and thought	2.40	.59	2.41	.09
Thought only	.53	-.57	3.86	4.01
Neither heard/thought	.37	-.95	.99	.86

Table J7

Pearson's r Correlation Matrix for Recognition Proportions for each Age and Test Condition Group

Group and Variable	1	2	3	4
Young – Standard Task				
1 Critical lure				
2 Weak lure	.47			
3 Unrelated item	.29	.83		
4 Presented item	.60	.67	.41	
Young – Source Task				
1 Critical lure				
2 Weak lure	.19			
3 Unrelated item	.29	.48		
4 Presented item	-.28	.52	.28	
Old – Standard Task				
1 Critical lure				
2 Weak lure	.50			
3 Unrelated item	.45	.77		
4 Presented item	.80	.71	.53	
Old – Source Task				
1 Critical lure				
2 Weak lure	.70			
3 Unrelated item	.36	.58		
4 Presented item	.69	.81	.45	

Table J8

Pearson's r Correlation Matrix for Source-Monitoring Response Options for each Age and Test Condition Group

	Young				Old			
Variable and Response	1	2	3	4	1	2	3	4
Critical lure								
1 Heard								
2 Heard and thought	.34				-.48			
3 Thought only	-.92	-.48			-.20	.05		
4 Neither heard/thought	-.04	-.20	-.28		-.72	-.06	-.34	
Weak lure								
1 Heard								
2 Heard and thought	.03				-.25			
3 Thought only	-.44	.25			.01	.18		
4 Neither heard/thought	-.32	-.72	-.58		-.69	-.27	-.64	
Unrelated item								
1 Heard								
2 Heard and thought	-.30				.26			
3 Thought only	-.15	.40			-.05	.07		
4 Neither heard/thought	-.33	-.45	-.85		-.60	-.37	-.75	
Presented item								
1 Heard								
2 Heard and thought	-.55				-.10			
3 Thought only	-.41	.40			.01	.32		
4 Neither heard/thought	-.44	-.41	-.39		-.90	-.25	-.39	

Table J9

Mauchly's Test of Sphericity for the Source-Monitoring Response Options for the Age Groups

	Young		Old	
Response Option Variable	χ^2	*p*	χ^2	*p*
Critical lure	42.84	<.01	17.31	<.01
Weak lure	14.85	.01	22.79	<.01
Unrelated item	32.68	<.01	48.44	<.01
Presented item	16.33	<.01	51.71	<.01

Note. For all calculations *df* = 5.

Appendix K: Chapter 6 ROC Curves

Overall False Alarm ROC Curves

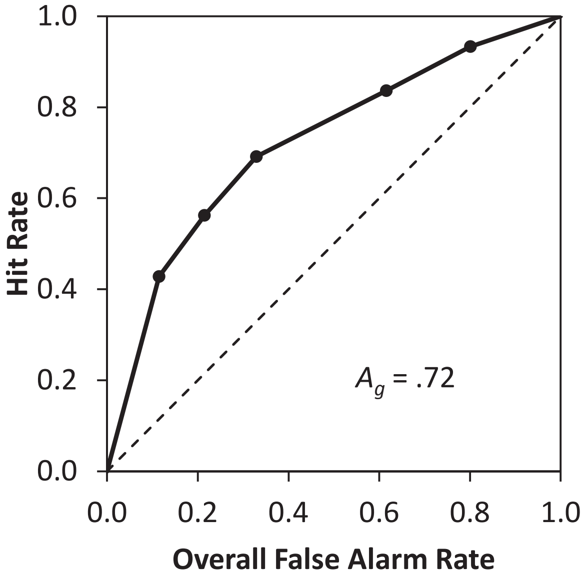


Figure K1. Younger adults in the standard condition ROC curve (based on overall false alarm rates).

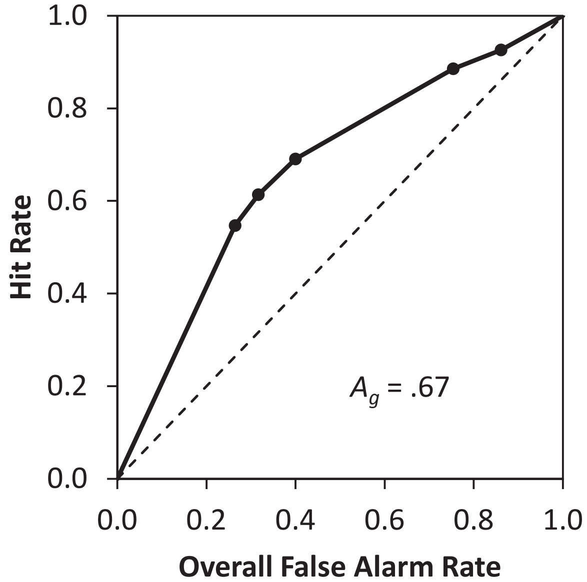


Figure K2. Older adults in the standard condition ROC curve (based on overall false alarm rates).

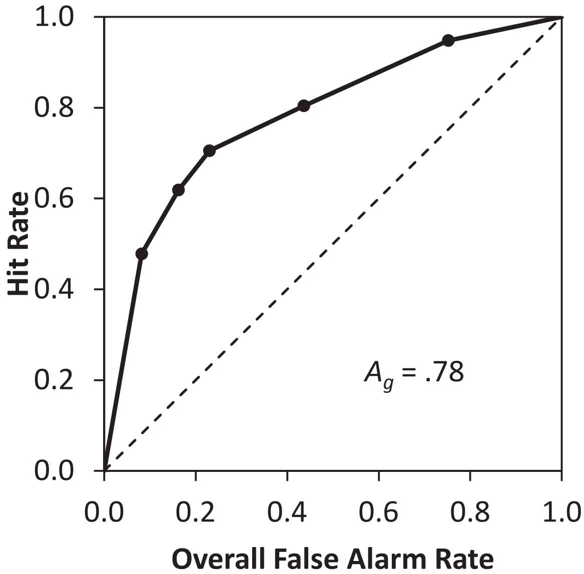


Figure K3. Younger adults in the source condition ROC curve (based on overall false alarm rates).

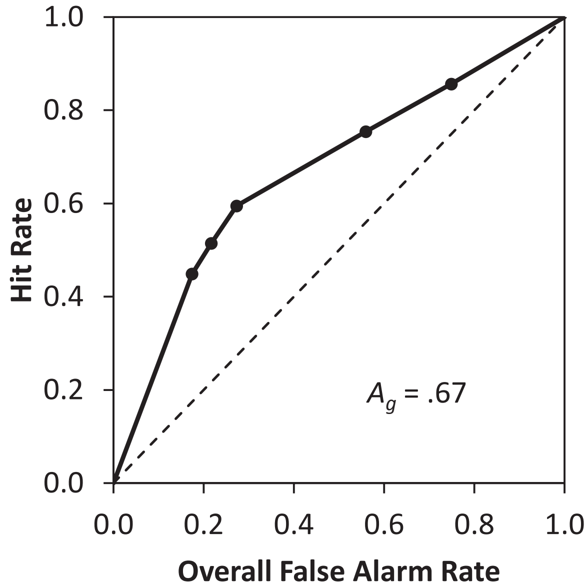


Figure K4. Older adults in the source condition ROC curve (based on overall false alarm rates).

Critical Lure False Alarm ROC Curves

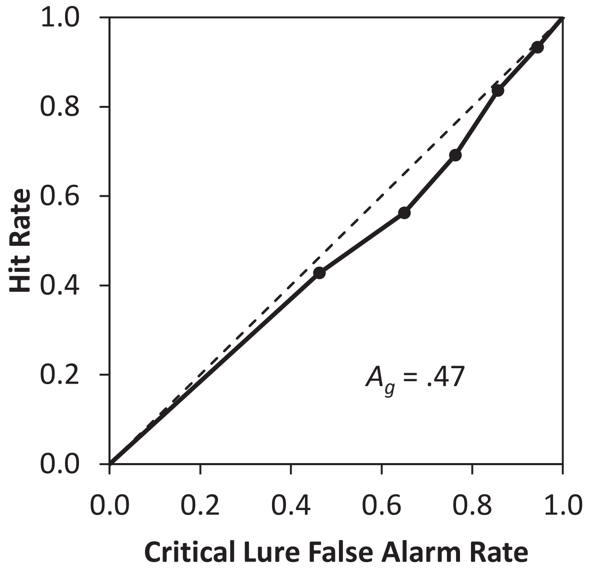


Figure K6. Younger adults in the standard condition ROC curve (based on critical lure false alarm rates).

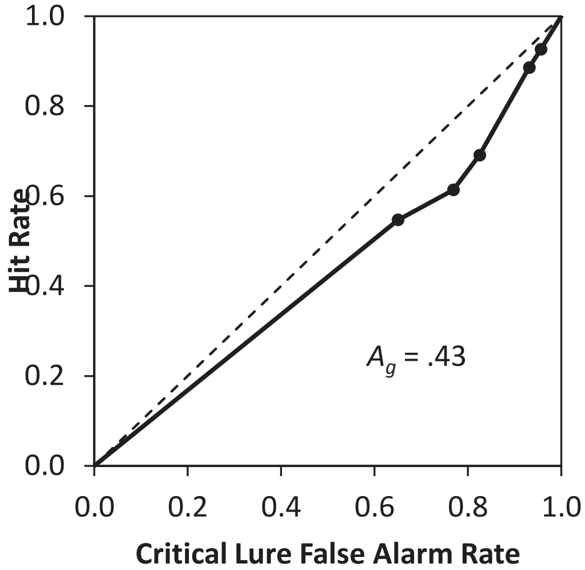


Figure K5. Older adults in the standard condition ROC curve (based on critical lure false alarm rates).

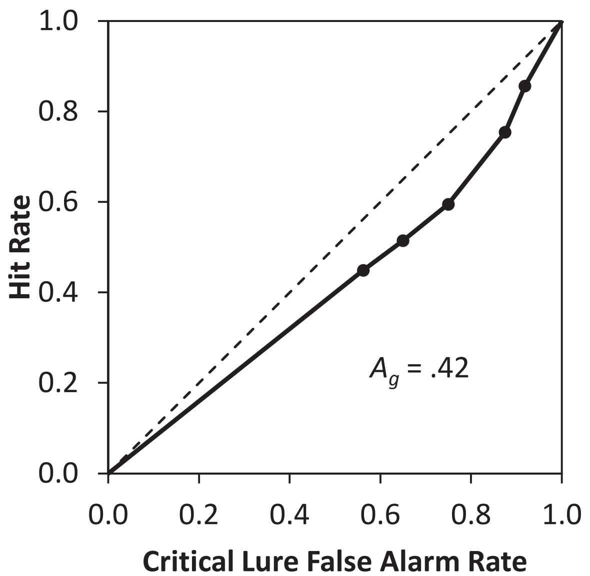


Figure K8. Younger adults in the source condition ROC curve (based on critical lure false alarm rates).

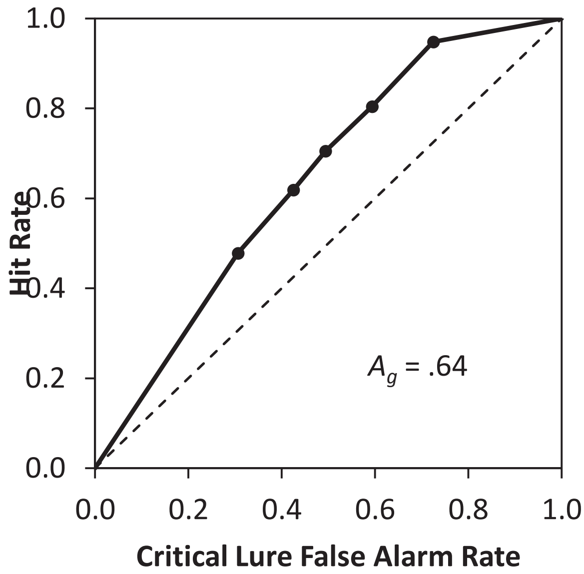


Figure K7. Older adults in the source condition ROC curve (based on critical lure false alarm rates).

Weak Lure False Alarm ROC Curves

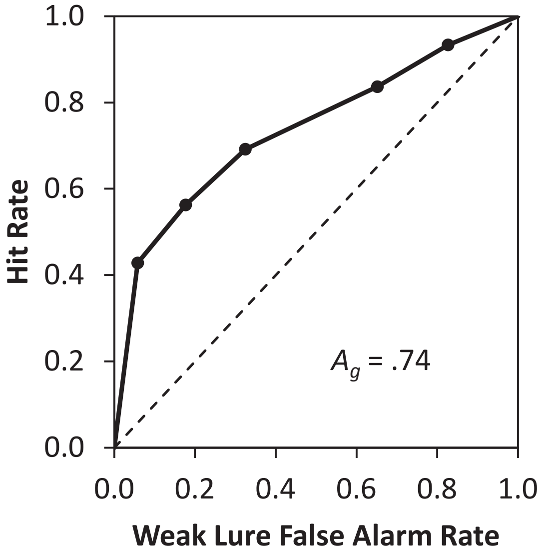


Figure K9. Younger adults in the standard condition ROC curve (based on weak lure false alarm rates).

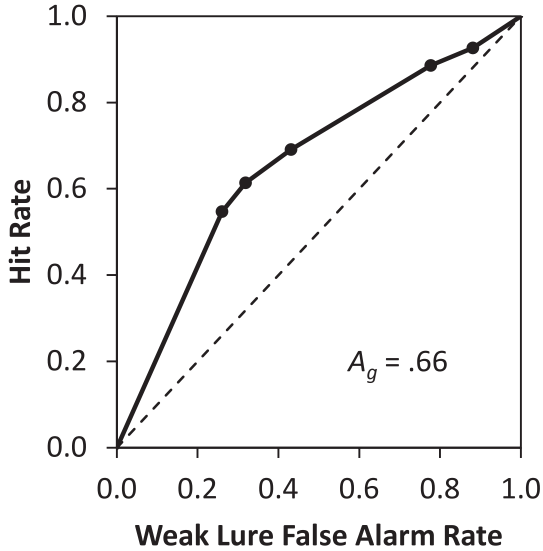


Figure K10. Older adults in the standard condition ROC curve (based on weak lure false alarm rates).

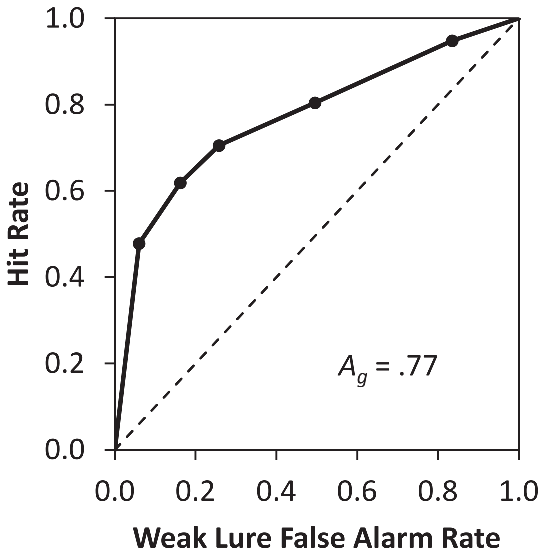


Figure K11. Younger adults in the source condition ROC curve (based on weak lure false alarm rates).

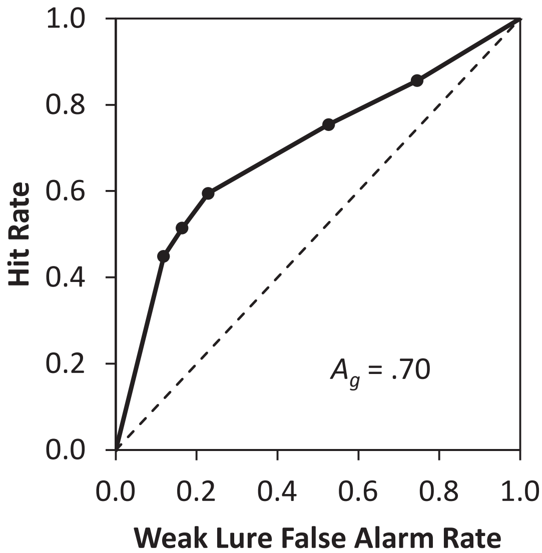


Figure K12. Older adults in the source condition ROC curve (based on weak lure false alarm rates).

Unrelated Item False Alarm ROC Curves

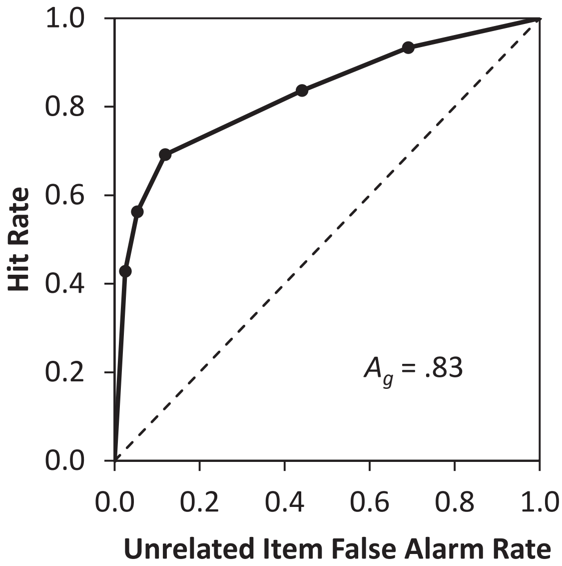


Figure K14. Younger adults in the standard condition ROC curve (based on unrelated item false alarm rates).

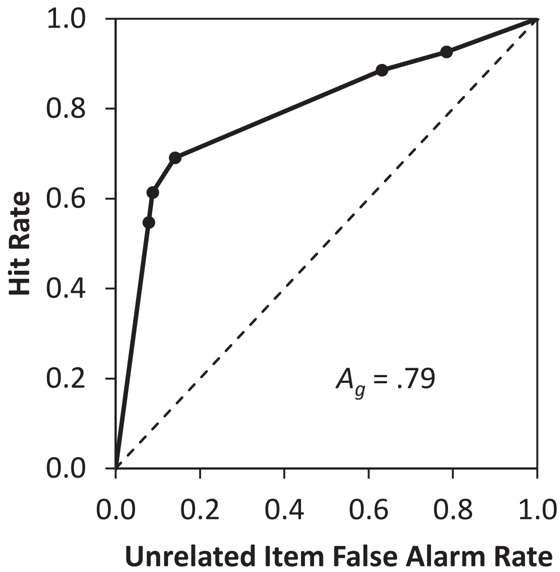


Figure K13. Older adults in the standard condition ROC curve (based on unrelated item false alarm rates).

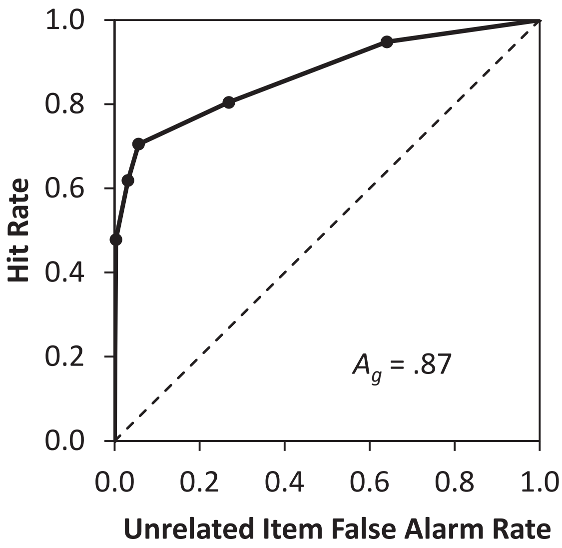


Figure K15. Younger adults in the source condition ROC curve (based on unrelated item false alarm rates).

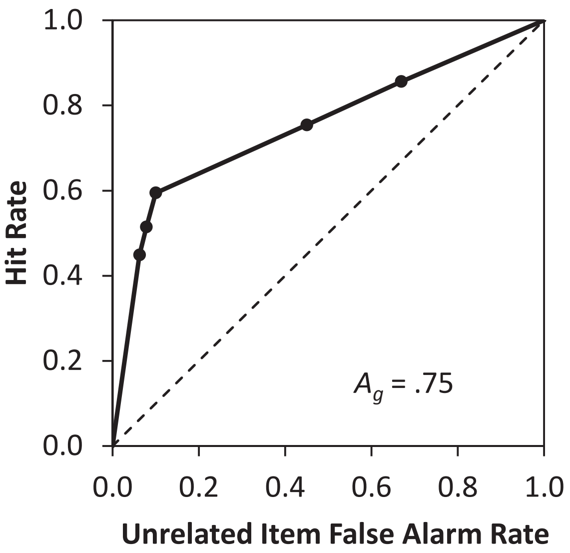


Figure K16. Older adults in the source condition ROC curve (based on unrelated item false alarm rates).

Appendix L: Chapter 6 *z*ROC Curves

Overall False Alarm *z*ROC Curves

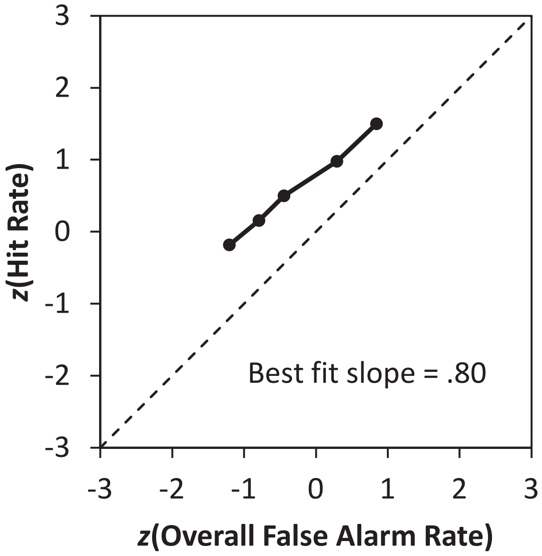


Figure L1. Younger adults in the standard condition *z*ROC curve (based on overall false alarm rates).

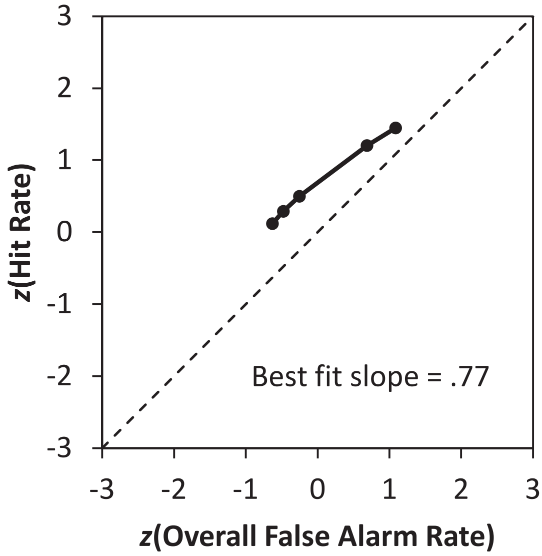


Figure L2. Older adults in the standard condition *z*ROC curve (based on overall false alarm rates).

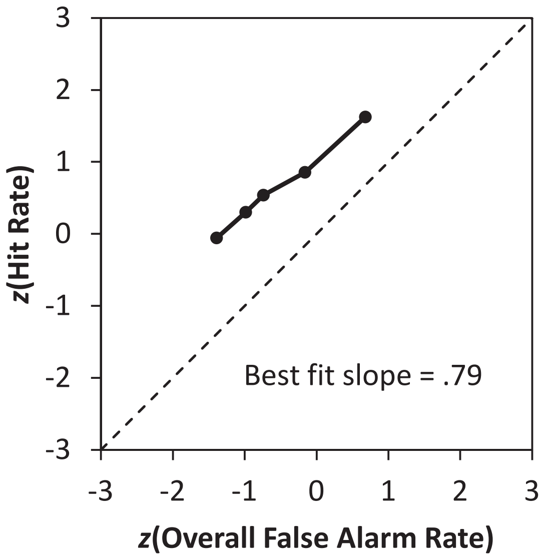


Figure L3. Younger adults in the source condition *z*ROC curve (based on overall false alarm rates).

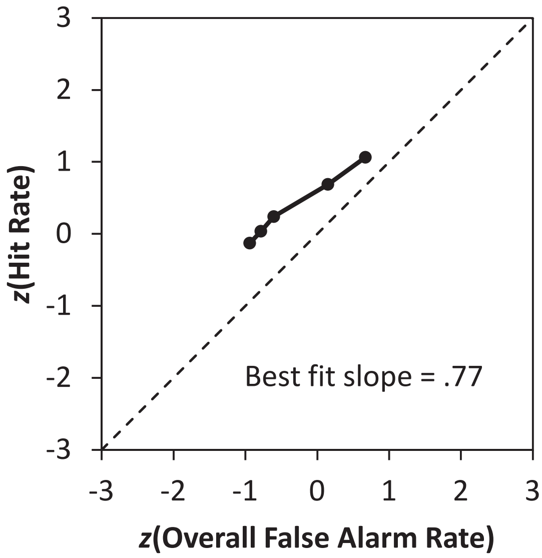


Figure L4. Older adults in the source condition *z*ROC curve (based on overall false alarm rates).

Critical Lure False Alarm *z*ROC Curves

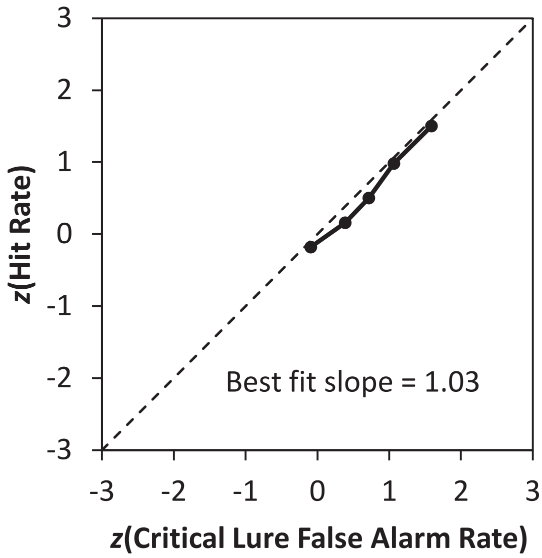


Figure L6. Younger adults in the standard condition *z*ROC curve (based on critical lure false alarm rates).

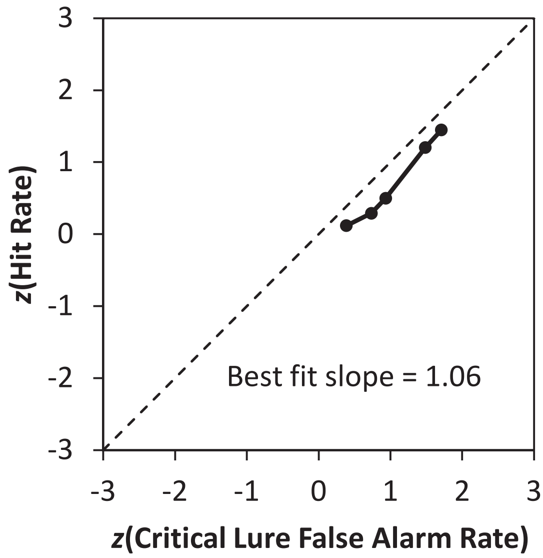


Figure L5. Older adults in the standard condition *z*ROC curve (based on critical lure false alarm rates).

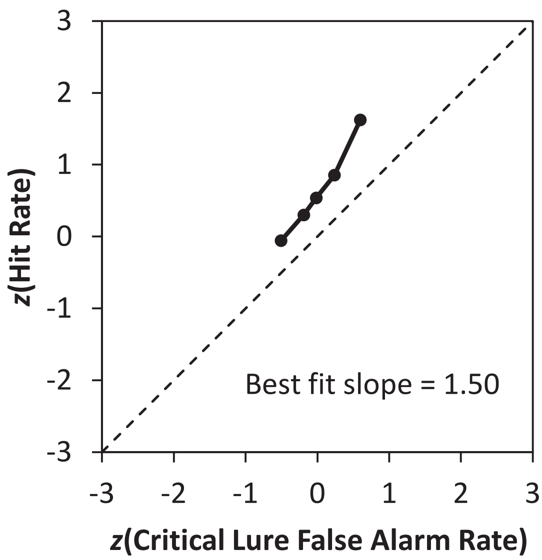


Figure L8. Younger adults in the source condition *z*ROC curve (based on critical lure false alarm rates).

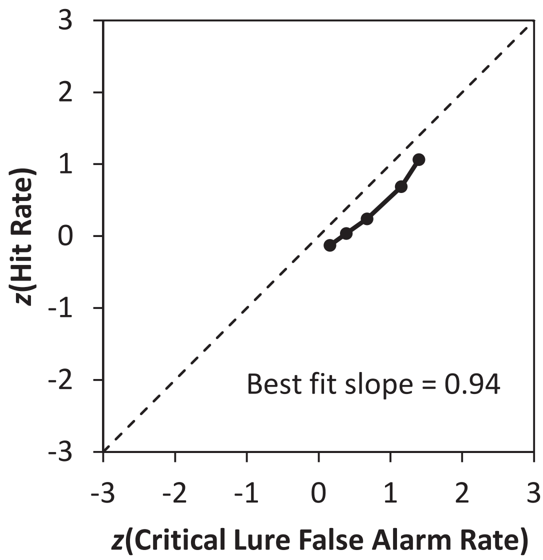


Figure L7. Older adults in the source condition *z*ROC curve (based on critical lure false alarm rates).

Weak Lure False Alarm *z*ROC Curves

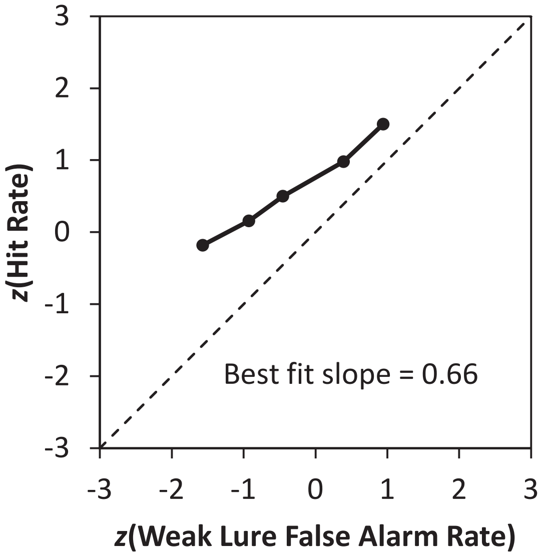


Figure L10. Younger adults in the standard condition *z*ROC curve (based on weak lure false alarm rates).

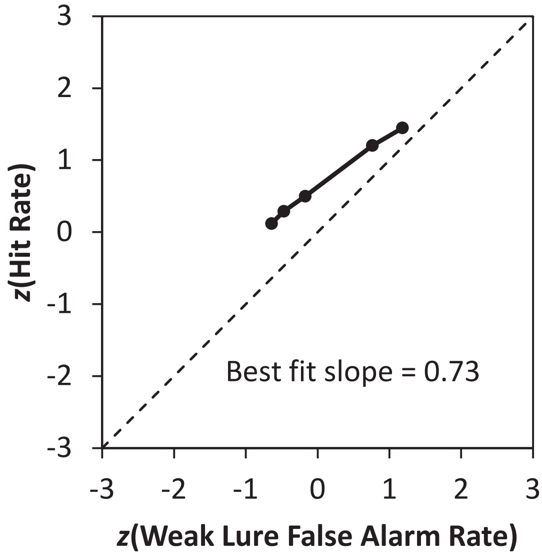


Figure L9. Older adults in the standard condition *z*ROC curve (based on weak lure false alarm rates).

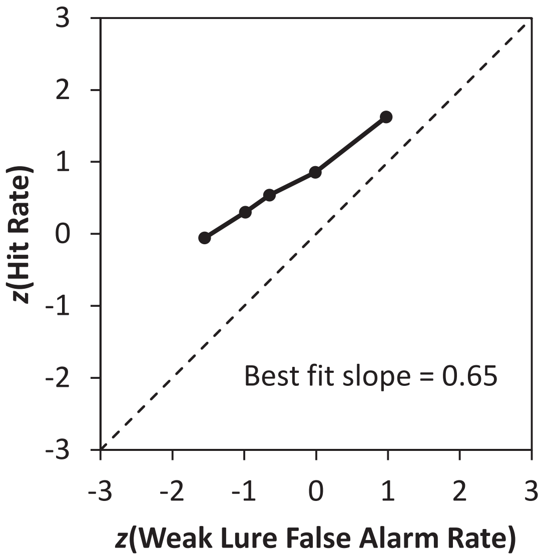


Figure L11. Younger adults in the source condition *z*ROC curve (based on weak lure false alarm rates).

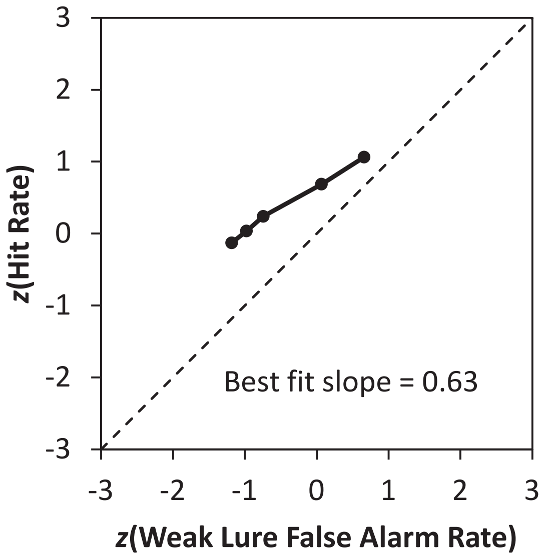


Figure L12. Older adults in the source condition *z*ROC curve (based on weak lure false alarm rates).

Unrelated Item False Alarm *z*ROC Curves

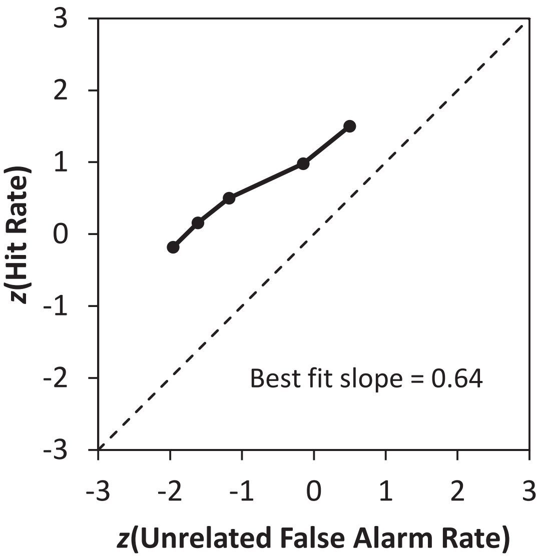


Figure L13. Younger adults in the standard condition *z*ROC curve (based on unrelated false alarm rates).

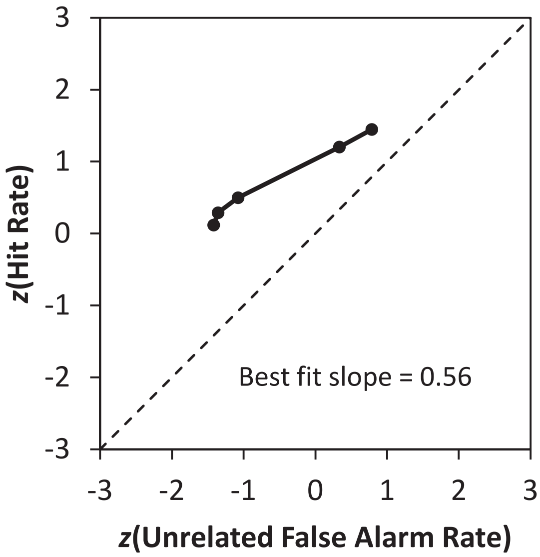


Figure L14. Older adults in the standard condition *z*ROC curve (based on unrelated false alarm rates).

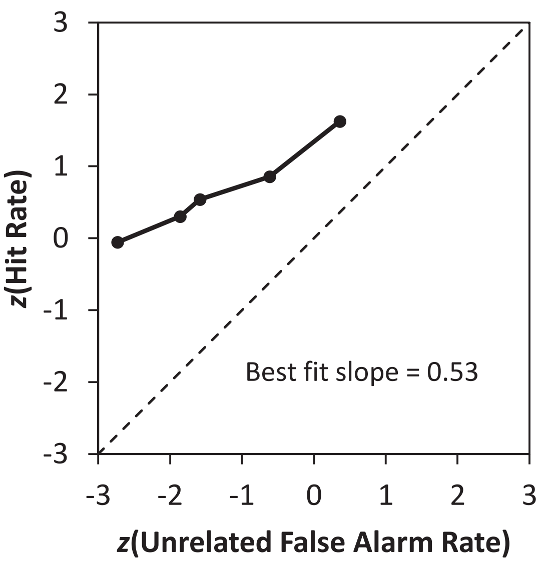


Figure L15. Younger adults in the source condition *z*ROC curve (based on unrelated false alarm rates).

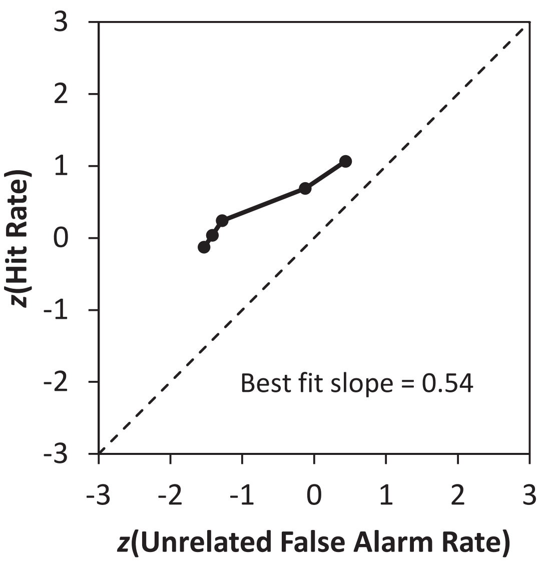


Figure L16. Older adults in the source condition *z*ROC curve (based on unrelated false alarm rates).

Appendix M: Chapter 6 Normality Test Results

Table M1

Kolmogorov-Smirnov Tests for the Overall False Alarm (FA) Rate and Corresponding d’ and c for the Age and Test Condition Groups

	Young				Old			
	Standard task		Source task		Standard task		Source task	
Variable	*D*	*p*	*D*	*p*	*D*	*p*	*D*	*p*
FA rate	.14	>.20	.11	>.20	.18	.10	.17	.13
d’	.14	>.20	.13	>.20	.13	>.20	.14	>.20
c	.08	>.20	.19	.06	.13	>.20	.16	.18

Note. For all calculations *df* = 20.

Table M2

Skewness and Kurtosis Z-scores for the Overall False Alarm (FA) Rate and Corresponding d’ and c for the Age and Condition Groups

	Young				Old			
	Standard task		Source task		Standard task		Source task	
Variable	z_{skewness}	z_{kurtosis}	z_{skewness}	z_{kurtosis}	z_{skewness}	z_{kurtosis}	z_{skewness}	z_{kurtosis}
FA rate	.10	-.58	1.57	1.37	1.61	1.03	1.45	.59
d’	-1.47	2.19	.41	-1.26	-.69	-.69	1.64	3.19
c	-.88	.01	-1.21	1.72	.04	.83	.07	-.25

Table M3

Kolmogorov-Smirnov Tests for the Separate d’ and c Indices for the Age and Test Condition Groups

	Young				Old			
	Standard task		Source task		Standard task		Source task	
Variable	*D*	*p*	*D*	*p*	*D*	*p*	*D*	*p*
Critical lure								
d’	.09	>.20	.14	>.20	.17	.12	.11	>.20
c	.17	.16	.13	>.20	.13	>.20	.09	>.20
Weak lure								
d’	.15	>.20	.13	>.20	.13	>.20	.14	>.20
c	.12	>.20	.17	.12	.12	>.20	.19	.06
Unrelated item								
d’	.21	.02	.18	.11	.15	>.20	.09	>.20
c	.20	.04	.23	.01	.13	>.20	.13	>.20

Note. For all calculations *df* = 20.

Table M4

Skewness and Kurtosis Z-scores for the Separate d' and c Indices for the Age and Test Condition Groups

	Young				Old			
	Standard task		Source task		Standard task		Source task	
Variable	$z_{skewness}$	$z_{kurtosis}$	$z_{skewness}$	$z_{kurtosis}$	$z_{skewness}$	$z_{kurtosis}$	$z_{skewness}$	$z_{kurtosis}$
Critical lure								
d'	-.76	-.05	.88	-.50	-1.82	.20	.18	-1.04
c	-1.03	1.04	.11	-.70	1.22	.42	-.12	-.75
Weak lure								
d'	.73	2.26	.24	-1.28	-1.16	.20	2.35	2.24
c	-.77	-.12	-1.06	.98	-.79	1.73	.40	-.66
Unrelated item								
d'	-2.07	.85	.69	-1.03	-1.81	2.49	.76	.25
c	-.29	-1.27	-1.96	.22	-.78	.35	-.83	.38