Memory function over time and across the lifespan:

Susceptibility to distortion and potential for modification

by

Kristin Eileen Flegal

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2011

Doctoral Committee:

Professor Patricia A. Reuter-Lorenz, Co-Chair
Associate Professor Cindy Lustig, Co-Chair
Associate Professor Linas A. Bieliauskas
Associate Professor Priti Shah

In memory of my dad,

who first taught me how to think like a scientist.

Acknowledgments

I would like to thank my advisors, Patti Reuter-Lorenz and Cindy Lustig, for their guidance and mentorship, the many privileged opportunities they have made available to me, the knowledge they have shared with me, and their unfailing support and encouragement throughout my graduate training. I am also truly grateful to my other committee members, Linas Bieliauskas and Priti Shah, for their expert opinions and valuable contributions to my dissertation.

I am indebted to the research assistants whose hard work made this research possible. In the Lustig lab, the assistance of Ridwaan Albeiruti, Molly Berenhaus, Michelle Cardinal, Katherine Cost, Lauren Davis, Magen Davis, Luke Gladstone, Elyssa Guslits, Jennifer Humfleet, Monica Kumar, Emily Kurzawa, Naomi Makofsky, Caitlin Mallory, Karen Mallozzi, Yosi Nacach, Phuoc Anh Nguyen, Jessica Oakley, Monika Potrykus, Kimberly Remski, Lauren Schiff, Gavin Stern, Surya Sabhapathy, Sasha Tereshchenko, Samantha Thomas, Maria Tocco, Taryn Weil, Madi Weisberg, and Mary Winters was invaluable for carrying out the training study in Chapter 3. In the Reuter-Lorenz lab, help from Tom Bergman, Greta Bielaczyc, Phebe Brenne, Ashley Brooks, Michael Dulas, Chelsey Forry, Autumn Gear, Elizabeth Irish, Alex Kapordelis, Michael Kinning, Sarah Koltun, Teija Madhusoodanan, Hiten Patel, Amanda Perez, Gabriela Ramirez, Emily Sallen, Puneet Sodhi, Fred Upton, Megan Walsh, Samantha

Winter, Laura Zahodne, and especially Halle Zucker, was instrumental to data collection for the experiments reported in Chapters 2 and 4. Elizabeth Irish, Teija Madhusoodanan, Hiten Patel, and Samantha Winter deserve special recognition for their help with preparing the references, tables, and appendices in this dissertation. I also wish to thank my fellow lab-mates, Katie Askren, Alex Atkins, Anne Berry, Kim Craig, Elise Demeter, Sara Festini, Ronit Greenberg, and Lynn Ossher, for their esteemed friendship as well as for the logistical support and insightful feedback on my work they have offered so frequently and so generously, without which I would have very little to show for my time in graduate school.

I am deeply grateful to my parents, Tom Flegal and Rita Kavanaugh, and my sisters, Tricia and Kerry, for the support and understanding that they have always given me. Most of all, I thank my husband Andrew for his love, patience, inspiration, and encouragement which has carried me through the last five years.

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

Human memory is imperfect, whether in failing to recall some piece of information from the past, or falsely remembering something that didn’t actually occur. Such errors are of special concern for older adults, as even healthy aging is associated with some decline in memory function. The present research investigates distortions within the putatively separate domains of short-term memory (STM) and long-term memory (LTM), and examines the potential for modifying memory performance through encoding strategy manipulations. In a false memory paradigm, the incidence of illusory recognition and accompanying measures of phenomenology (ratings of confidence and recollection) were equivalent under STM and LTM conditions. A levels-of-processing manipulation demonstrated that deep encoding (relative to shallow encoding) was associated with increased false memory rates emerging at long-term testing. Relational encoding, relative to item-specific encoding, was associated with higher levels of recollective phenomenology, even at short delays. Predicted reductions in false memory under item-specific encoding instructions were not observed. In both encoding manipulation experiments, false memory phenomenology was found to be stable from STM to LTM. In a memory training study with older adults, environmental support in the form of fixed encoding time produced training-task improvements as large as those resulting from instruction with an experimenter-provided relational encoding strategy. Furthermore, on a post-training false memory test, only participants who followed the relational encoding strategy did not show a correlation between better training-task performance and greater resistance to false memory, suggesting that the benefit to accurate memory came with the cost of increased vulnerability to related but misleading information. Taken together, these studies shed light on the outcomes of processes engaged at encoding, and reveal commonalities favoring the operation of unitary memory processes at short and long delays.

CHAPTER 1

INTRODUCTION

The research presented in this dissertation has been motivated by the goal of understanding the mechanisms of accurate and inaccurate remembering, with a particular focus on the outcomes of processes engaged at encoding. The following chapters report on a series of experiments that I have conducted to investigate two general features of human memory: its susceptibility to distortion, and its potential for modification. I will begin with a brief review of the relevant background literature on memory errors, memory systems, memory processes, memory aging, and memory training.

Memory distortions and illusions

Human memory is fallible, and it fails to varying degrees. For example, a person may fail to remember some aspects of an experience altogether, or remember some aspects of an experience inaccurately, or remember aspects of an experience that never happened at all. The inherently reconstructive, rather than reproductive, nature of memory has been exposed in numerous classic studies from cognitive psychology, from Bartlett's findings on the contributions of pre-existing knowledge and expectations to memory errors (Bartlett, 1932), to

Loftus’ renowned experiments on the influence of post-event misinformation in determining how accurately an event will be later remembered (Loftus, Miller, & Burns, 1978; Loftus & Palmer, 1974).

Contemporary research on memory distortions has capitalized on the popularity of a laboratory task known as the converging associates or DRM paradigm (Deese, 1959; Roediger & McDermott, 1995), which was developed to probe false memories of experiences that are associated with or otherwise highly similar to actual events. In this procedure, participants study lists of semantically-related words (e.g., *bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, nap, peace, yawn, drowsy*) which are all associates of an unstudied theme word, the “critical lure” word for each list (in this example, *sleep*). Semantic memory distortions are evident in the critical lure words being both falsely recalled (typically as frequently as are studied words from the middle of the serial position curve) and falsely recognized (often at a frequency approaching the hit rate for studied words) with remarkable regularity; moreover, false memories of these lure items are often accompanied by ratings of high confidence or strong feelings of subjective recollection (e.g., Anastasi, Rhodes, & Burns, 2000; Lampinen, Neuschatz, & Payne, 1998; Roediger & McDermott, 1995). These memory illusions are well established in the domain of episodic, long-term memory (Brainerd & Reyna, 2005; Gallo, 2006), but recent studies have examined the question of whether the processes responsible are in fact unique to long-term memory tasks.

Remembering (and misremembering) over the short-term and long-term

Memory research has long been guided by a distinction between systems subserving the temporary retention of information (short-term memory; STM) and those underlying a permanent store of knowledge and experiences (long-term memory; LTM). Evidence from behavioral observations (Baddeley, 1966; Glanzer & Cunitz, 1966), from specific memory deficits exhibited by patients who have suffered neurological trauma (Scoville & Milner, 1957; Warrington & Shallice, 1969), and from early neuroimaging studies (D'Esposito, Postle, & Rypma, 2000; Schacter & Wagner, 1999) favored models of multiple memory systems by emphasizing the functional and anatomical separability of short-term and long-term mnemonic processes.

However, further research from cognitive psychology, cognitive neuropsychology, and cognitive neuroscience has suggested that STM and LTM primarily rely on common, rather than separate, component processes and neural substrates (Jonides et al., 2008; Ranganath & Blumenfeld, 2005). Recent demonstrations of semantic memory distortions occurring within the capacity and retention limits of canonical STM task conditions (Atkins & Reuter-Lorenz, 2008; Coane, McBride, Raulerson, & Jordan, 2007) also challenge the dissociability of STM and LTM systems purported by traditional memory models. One way of testing the alternative hypothesis, that a unitary system is responsible for memory function at short and long delays, is to evaluate the influence of encoding manipulations under STM and LTM conditions.

Encoding influences

Much research on learning and memory has explored the memorial consequences of how information is processed, or encoded, when it is initially encountered. For example, Craik and Lockhart (1972) proposed a levels-of-processing theory to argue that memory durability has more to do with the “depth of analysis” that items receive rather than their presence in one store or another (e.g., STM or LTM). According to this viewpoint, better memorability is associated with information processing at a *deep*, semantic level in which analysis is meaning-based, compared to a *shallow*, perceptual level in which analysis is surface-based. Support for these claims came from evidence that superior memory resulted when the meanings of words were judged as fitting into the context of a sentence appropriately or not compared to when the sounds of words were judged as rhyming or not, for example (Craik & Tulving, 1975). Another theoretical perspective on the influence of encoding contrasts *item-specific* processing, which draws attention to characteristics that differentiate studied items, with *relational* processing, which draws attention to characteristics that studied items have in common (Einstein & Hunt, 1980; Hunt & Einstein, 1981). While both item-specific and relational processing can be considered forms of “deep” encoding, they are selectively advantageous to later memory performance depending on whether the material to be learned is intrinsically related (e.g., lists of semantic associates), in which case item-specific processing will be most beneficial, or intrinsically unrelated, in which case relational processing will be most beneficial (Hunt & McDaniel, 1993).

Some of the first studies of encoding influences in the DRM literature used levels-of-processing manipulations, and showed that deep, meaning-based encoding increases not only accurate memory, but also false memory for unstudied but related material (Rhodes & Anastasi, 2000; Thapar & McDermott, 2001; Toglia, Neuschatz, & Goodwin, 1999). “Fuzzy trace” theory (Brainerd & Reyna, 2002; Reyna & Brainerd, 1995) proposes that two types of information are encoded into memory in parallel: verbatim, item-specific traces which faithfully record surface features and details of an experience, and gist representations which store general themes or semantic content. The strength of the two traces diverges over time, because the integrity of verbatim traces does not last long without active refreshing. According to this account, the results from levels-of-processing DRM experiments can be explained in that deep encoding strengthens gist traces, which support both accurate memory and false memory for semantically related lists, and shallow encoding strengthens verbatim traces, which can oppose false memory only as long as they are available.

Item-specific processing has been proposed as an alternate encoding strategy that may bestow a similar benefit to accurate memory as does deep encoding, but without the cost of also increasing false memory (Arndt & Reder, 2003; Hege & Dodson, 2004; McCabe, Presmanes, Robertson, & Smith, 2004). Because relational processing is redundant with the intrinsically related structure of DRM word lists, it is likely to amplify the encoding of gist information, thus increasing vulnerability to false memory, while item-specific processing is likely to counteract that tendency by suppressing the encoding of gist information.

Impairments in item-specific processing, and a concomitant over-reliance on gist information, have been suggested to contribute to inflated false memory rates observed in some patient populations (Budson, Daffner, Desikan, & Schacter, 2000) and in older adults (Butler, McDaniel, McCabe, & Dornburg, 2010; Kensinger & Schacter, 1999; Tun, Wingfield, Rosen, & Blanchard, 1998).

Aging and memory distortions

Even in healthy aging, later years of life bring changes in memory function. Compared to young adults, older adults experience greater difficulties with source memory, which involves recovering specific details about the context in which information was initially encountered (Glisky, Rubin, & Davidson, 2001; Hashtroudi, Chrosniak, & Johnson, 1990; Schacter, Kaszniak, Kihlstrom, & Valdiserri, 1991), and rely to a greater extent on familiarity than recollection in their memory decisions (Jacoby, 1999) – an expected consequence of deficits in encoding or retrieving diagnostic source information. Research using the DRM paradigm has shown that aging is also associated with relative increases in false memory compared to accurate memory (Balota et al., 1999; Norman & Schacter, 1997; Schacter, Koutstaal, & Norman, 1997). This increased susceptibility to memory distortions has been suggested to reflect the combination of an age-related decline in source memory with unchanged semantic associative processes (Dehon & Bredart, 2004).

While source memory impairments and a preference for familiarity-based processing may predispose older adults to rely more on gist information (Dennis, Kim, & Cabeza, 2007; Koutstaal & Schacter, 1997), an important finding in the DRM literature has been that older adults can benefit from manipulations designed to decrease false memories (Dodson, Koutstaal, & Schacter, 2000; Koutstaal, Schacter, Galluccio, & Stofer, 1999). These include studies that limit gist-based processing at encoding (Thomas & Sommers, 2005) or provide instruction in the use of a "distinctiveness heuristic" to improve memory decisions at retrieval (Dodson & Schacter, 2002a; Schacter, Israel, & Racine, 1999). Whether by aiming to reduce false memories, as in these examples, or to boost accurate memories, research on the potential for modifying memory function has important real-world applications – especially for enhancing the cognitive fitness of older adults.

Interventions to improve memory accuracy

Memory training interventions have received a great deal of attention in the cognitive aging literature. In addition to being more susceptible to false memories than young adults (Balota, et al., 1999; Schacter, et al., 1997), older adults often fail to self-initiate the deep, associative encoding processes that facilitate later accurate memory (Braver et al., 2001; Craik & Byrd, 1982; Paxton, Barch, Racine, & Braver, 2008). However, training studies have established that older adults can benefit from instructions and environmental support to encourage such processing (e.g.,Logan, Sanders, Snyder, Morris, & Buckner, 2002; Naveh-Benjamin, Brav, & Levy, 2007; Paxton, Barch, Storandt, & Braver, 2006) and change their memory performance to more closely resemble young adults'. This is a theoretically important finding which indicates that effortful cognitive abilities are not entirely lost with age, but are frequently latent in the absence of external support. Discovering the most effective ways of targeting intact but inefficiently used processes for training is a promising direction for applied memory research with older adults.

Training older adults to use experimenter-provided mnemonic strategies is one approach to improving memory accuracy, and indeed this is often effective for increasing performance on the task being trained (e.g., Ball et al., 2002; Kliegl, Smith, & Baltes, 1989; Rebok & Balcerak, 1989). However, this approach rarely produces transfer, or generalization, of training benefits to other tasks (Lustig, Shah, Seidler, & Reuter-Lorenz, 2009), and another significant problem is that individual differences in existing cognitive strengths and weaknesses will limit the usefulness of any particular memory strategy for a representative sample of older adults (Hill, Yesavage, Sheikh, & Friedman, 1989; Verhaeghen & Marcoen, 1996; Yesavage, Sheikh, Tanke, & Hill, 1988). An alternative approach is to provide environmental support for self-initiated processing but no explicit mnemonic strategy, encouraging older adults to generate their own strategies tailored to their personal preferences and pre-existing strengths. The findings from a small number of published studies examining older adults' self-generated memory strategies are that in general they are at least as effective as experimenter-provided mnemonics (Baltes, Sowarka, & Kliegl, 1989; Derwinger, Stigsdotter Neely, Persson, Hill, & Bäckman, 2003; Hill, Allen, & Gregory, 1990) and may produce long-lasting benefits which are less dependent on external support (Derwinger, Stigsdotter Neely, & Bäckman, 2005).

Overview of the present dissertation

The present research addresses the question of how memory accuracy can be enhanced, by investigating factors that govern its strengths and weaknesses. The motive of better understanding the origins of distortions and failures of accurate memory – and leveraging that knowledge to improve memory accuracy through behavioral interventions – links the three components of this dissertation, summarized below.

Chapter 2 presents findings from a pair of experiments which show surprisingly delay-invariant false memory effects and phenomenology in STM and LTM. In a modified version of the DRM procedure, participants studied lists of four semantically-related words and were probed immediately following a filled 3-4 second retention interval or approximately 20 minutes later in a surprise recognition test. A false memory effect is expressed in higher rates of false alarms to unstudied but semantically-related probes than to unstudied, unrelated probes. Results from this study revealed that the incidence of false recognition, along with the confidence ratings and Remember/Know judgments assigned to false memories, were equivalent under STM and LTM conditions. These findings challenge conventional models of multiple memory systems which characterize semantic coding as the province of LTM only, suggesting instead that the processes responsible for semantic memory errors are relatively time-invariant.

Chapter 3 reports outcomes from a memory training intervention for older adults. Focusing on LTM, this study examined training and transfer effects in conditions that either 1) mandated a deep, associative encoding strategy, 2) attempted to suppress such encoding by mandating rote rehearsal, or 3) encouraged effort towards encoding (by enforcing study times) but allowed participants to choose their own strategies. Participants completed seven 1-hour sessions of training scheduled over the course of three weeks. They also completed a battery of other, untrained tasks (used to measure the transfer of training to other situations) before and after training. Results from this study showed that environmental support to facilitate the self-initiation of effortful memory processes diminished the association between older age and poorer training task performance that was present under open-ended conditions in an earlier experiment (Bissig & Lustig, 2007). Furthermore, results from a DRM task suggested that training with a deep, associative encoding strategy may carry the cost of decreased resistance to false memory for related but misleading information, in addition to the benefit of higher accurate memory for relevant information.

Chapter 4 presents results from a further two experiments which integrated the novel experimental procedure introduced in Chapter 2 (directly comparing false memories in the same participants under STM and LTM conditions) with the encoding manipulations investigated in Chapter 3. Participants studied lists of four semantically-related words, under varying encoding strategy instructions, and were probed immediately following a filled 3-4 second retention interval or approximately 20 minutes later in a surprise recognition test. In a levels-of-processing manipulation, deep encoding (relative to shallow encoding) preserved accurate memory across delay and increased false memory at LTM, with minimal effects at STM. Collapsed across encoding condition, the incidence and phenomenological qualities of false memory did not differ between short-term and long-term testing, replicating the findings reported in Chapter 2. In a separate experiment comparing item-specific and relational processing, relatively time-invariant effects were observed in measurements of false memory phenomenology, but other results did not support hypotheses based on theoretical considerations and results from previous DRM studies.

Taken together, the work contained in this dissertation aims to make a contribution at the intersection of experimental and applied research on memory. The three studies reported in the following chapters share a focus on better understanding the mechanisms that influence memory accuracy, with the hope that their findings can help to guide the subsequent development of effective memory training interventions, and advance our knowledge of how we remember and why we forget.

CHAPTER 2

FALSE MEMORIES SECONDS LATER: THE RAPID AND COMPELLING ONSET OF ILLUSORY RECOGNITION

Introduction

False memories refer to distortions of the source, details, or meaning of past experiences. In the laboratory, false long-term memories (LTM) are reliably produced with the converging associates or DRM paradigm (Deese, 1959; Roediger & McDermott, 1995), in which unstudied semantic associates (“related lure” words) are misremembered as studied items (see Gallo, 2006, for a comprehensive review). Moreover, false memories can be accompanied by high confidence ratings or strong feelings of recollection. Participants in DRM studies will avow qualitative memories of having studied lure words, and may feel equally confident in their recognition of lure and studied words (e.g., Anastasi, et al., 2000; Roediger & McDermott, 1995). Theoretical explanations of false memory are typically framed in terms of LTM processes, invoking associative activation and source monitoring failures (Gallo & Roediger, 2002; Robinson & Roediger, 1997; Roediger, McDermott, & Robinson, 1998), and further implicating variable strengths of verbatim memory traces and gist memory for semantic content (Brainerd, Yang, Reyna, Howe, & Mills, 2008; Reyna & Brainerd, 1995).

Recently, however, false memory effects have been reported in the domain of short-term memory (STM). Findings of false recognition and false recall from lists of four semantically-related words occurring after mere seconds (Atkins & Reuter-Lorenz, 2008; Coane, et al., 2007) suggest that the processes responsible for false memories may be relatively delay-invariant. The vulnerability of STM to such distortion contrasts with conventional models of multiple memory systems which characterize semantic codes as the province of LTM (Baddeley, 1986), but converges with evidence for a unitary memory system with delay-invariant storage and retrieval processes (Jonides, et al., 2008). Testing the dissociability of false memories at short and long delays is thus pertinent to elucidating the architecture of memory.

The present report directly compares false memories in the same participants under STM and LTM conditions. We used a novel experimental procedure that minimizes encoding differences and honors the limited capacity and brief storage demands of canonical STM tasks, while exploiting the large capacity and longer delays that characterize LTM. Unlike previous DRM studies (using lists of 12-15 items) that manipulated retention interval (e.g., Colbert & McBride, 2007; Lampinen & Schwartz, 2000; Seamon et al., 2002; Thapar & McDermott, 2001), our approach probes four-item lists individually after 3-4 seconds, or with continuous recognition after all lists are encoded (on average 20 minutes later). Therefore, we can compare both quantity (the relative incidence of false recognition) and quality (the accompanying phenomenology) of memory distortions in the context of two putatively separate memory domains.

Several considerations predict semantic distortions to be more prevalent and compelling in LTM than STM. First, proactive and retroactive interference in LTM should contribute to greater decline of verbatim traces compared to STM, where retroactive interference, in particular, is minimal. Thus verbatim memory, used to monitor and oppose false memory, should have an advantage in STM. Second, context and source information are salient in the short term because each probe is assessed exclusively in relation to the immediately prior memory set. Consequently, STM conditions provide more cues for monitoring recognition decisions compared to LTM. Third, according to traditional models of separable memory systems, semantic coding that engenders false memories in the DRM paradigm should predominate perceptual coding in LTM, whereas the opposite prevails in STM. Relatedly, despite recent imaging evidence challenging separable memory systems (Ranganath & Blumenfeld, 2005), studies continue to find more robust meaning-based encoding effects in LTM than STM (Rose, Myerson, Roediger, & Hale, 2010). At minimum, semantic processing might be expected to exert measurably greater influence under LTM than STM conditions.

Similar considerations predict the phenomenology of false memories to differ in the short versus long term. The availability of verbatim traces in STM should permit diagnostic monitoring with the relative absence of detail-rich memories for lure probes supporting their correct rejection, or at minimum their endorsement with low confidence or a context-free sense of familiarity. If memory distortions are not found to be more prevalent or compelling in LTM than STM, however, process constancy would be implicated across memory domains.

As the results from two experiments reported here indicate, dissociations predicted by models of multiple memory systems were not observed. The absolute incidence of false memories was greater under LTM than STM conditions, but when corrected for the increased frequency of false alarms to unrelated probes, false recognition rates did not significantly differ. Moreover, our findings indicate the phenomenology of false memory errors is equally compelling in STM or LTM trials, suggesting a rapid breakdown of monitoring processes.

Experiment 1

Method

Participants

27 individuals (18-20 yrs old) participated for course credit. Three participants were excluded for recognition accuracy scores > 2.5 standard deviations from the mean. Research protocols were approved by the University of Michigan Institutional Review Board, and all participants provided written informed consent.

Materials

The memory sets were 128 lists of four semantically-related words, all associates of a common theme word. Forty-eight lists were created from published lists used in the DRM paradigm (Roediger, Watson, McDermott, & Gallo, 2001), as previously described (Atkins & Reuter-Lorenz, 2008). Eighty new lists were constructed by selecting theme words from the University of South

Florida (USF) Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) with the following constraints: Each new theme word (1) had frequency and concreteness values within the range of the original 55 DRM lists; (2) had at least four semantic associates; (3) shared no associates with any other theme word. List items were cross-checked across associative matrices from the USF norms to minimize semantic associations between lists. Pilot testing confirmed that the new lists induced false recognition in the Atkins and Reuter-Lorenz (2008) STM task.

The 128 lists were divided into four groups of 32 four-word lists (Groups A-D) equated in mean backward associative strength ($M = 0.35$), and containing an even distribution of old (i.e., derived from DRM) and new lists. Each group was further divided into two subgroups of 16 four-word lists, following the same parameters, to balance the status of each list as short-term versus long-term memoranda across participants.

The four semantically-related words in each list converged upon a theme word (e.g., SLEEP for the associates nap, doze, bed, awake) which served as the probe on all trials. The three probe types were: *related lure*, the unstudied theme word associated with a studied list; *unrelated lure*, an unstudied theme word associated with a nonpresented list; and *target*, the theme word associated with, and present in, a studied list (replacing one of the four associates; ordinal position of the replaced item balanced across trials). Probe type was counterbalanced with word lists across participants, so that for one quarter of all participants, lists in Group A were paired with related lures, B with unrelated lures, and C with target probes. Theme words associated with nonpresented Group D lists served as the unrelated lure probes. For half of the participants in each of these four counterbalanced orders, the first subgroup of lists from each group (e.g., A1) was probed during the STM trials, and the second subgroup of lists (e.g., A2) during the LTM trials; the assignment was reversed for the other half of the participants. This procedure ensured that all participants encountered the same probes – all theme words, but in different contexts – as related lure, unrelated lure, or target probes, and as STM or LTM probes. Theme words appeared only once during the experiment; no list was probed in both STM and LTM trials.

Design and procedure

The STM task used by Atkins and Reuter-Lorenz (2008) was modified to test recognition at both short and long delays (see Figure 2.1). Four-word lists were probed either within the same trial (i.e., STM) or in a surprise recognition test following completion of all STM trials (i.e., LTM), to compare performance across equivalent lists without confounding long-term recognition with intervening short-term recognition of the same lists. STM trial parameters were those used by Atkins and Reuter-Lorenz, with two exceptions: (1) a confidence rating followed the response to every probe word; (2) two trial types were randomly interspersed: half of the trials ended with a probe word and confidence rating, and the other half ended with cues for arbitrary button presses with corresponding response mappings; these memory sets were subsequently probed at LTM.

STM trials. Each STM trial began with a four-item memory set presented for 1,200 msec. During the 3,000-4,000 msec retention interval, participants completed a math verification problem based on the operation span task (Turner & Engle, 1989) by making a left-handed response to indicate whether or not an equation was solved correctly[1]. Next, the probe word appeared for 3,000 msec, and participants indicated whether it was in the memory set with a left-handed response (*yes* or *no*). Lastly, a four-point confidence scale appeared for 3,000 msec, and participants used their right hand to indicate their confidence in the response they just made: *very low, somewhat low, somewhat high,* or *very high*. On trials not probed at STM, a display of two boxes replaced the probe word, to prompt an arbitrary left-handed response, and then a display of four boxes replaced the confidence scale, to prompt an arbitrary right-handed response. Each participant completed 96 STM trials presented in random order. Of the 48 trials probed at STM, 16 were of each probe type (related lure, unrelated lure, target).

LTM trials. A 2-min break followed completion of the STM trials, then participants were informed about, and given instructions for, the LTM recognition test. Each participant completed 96 LTM trials, 48 of which tested the memory sets that were not probed at STM (16 of each probe type). Additionally, there were 18 trials of studied associates from memory sets that were probed at STM (never including theme words from target probe trials), and 30 trials of unstudied,

---

[1] The filled retention interval parallels the additional "processing" demands that characterize working memory measures (e.g., Turner & Engle, 1989). Atkins and Reuter-Lorenz (2008) tested both filled and unfilled intervals in the STM paradigm and verified that false memory errors occurred reliably in both conditions. In the present paper we use the terms short-term memory and working memory interchangeably.

unrelated foils, matched for frequency and word length with the corpus of theme words used in the experiment.

In each LTM trial, a probe word first appeared for 3,000 msec, and participants made a left-handed response to indicate whether or not it was present during the STM trials. A four-point confidence scale (as before) then appeared for 3,000 msec, and participants used their right hand to indicate their confidence in their preceding response.

Trials without responses, with non-allowable responses or with response times under 200 msec were excluded from all subsequent analyses. Trimming retained 93% of STM trials and 99% of LTM trials, and did not change the pattern of results.

Results and Discussion

To assess the effects of probe type and delay on recognition accuracy and confidence, we conducted 2×2 within-subjects ANOVAs, and paired *t* tests for subsequent analyses. Effect sizes were computed using original standard deviations for each condition mean (Dunlap, Cortina, Vaslow, & Burke, 1996). Unless otherwise stated, statistical tests are significant at $p < .01$.

Accuracy

Mean math task accuracy during the STM trial retention interval was 0.89. The key comparison between *yes* responses to unstudied, related lure probes and unstudied, unrelated lure probes revealed a false memory effect at both STM ($t(23) = 4.81$, $d = 1.43$) and LTM ($t(23) = 5.13$, $d = 1.07$; Table 2.1A). The lure type (related, unrelated) × delay (STM, LTM) interaction was not significant ($F(1,23) = 1.08$, $p = .31$, $\eta_p^2 = .05$). Surprisingly, this result indicates that the rate of false recognition did not increase from short-term to long-term testing. Thus, while false alarms overall were more frequent for LTM than STM, as veridical memory dropped over time, corrected false memory rates (i.e., corrected rates of false alarms to related lures) were statistically indistinguishable in the short and long term. At LTM, the related lure false alarm rate approached, but remained reliably lower than, the hit rate for target probes ($t(23) = -3.82$, $d = .72$).

We used two approaches to obtain comparable measures of memory performance across delay (see also Colbert & McBride, 2007; Seamon, et al., 2002): (1) true and false recognition accuracy conditionalized by subtracting the baseline (unrelated lure) false alarm rate from target hits and related lure "hits" (producing the discriminability index *Pr*), and (2) estimates of recognition sensitivity (d') computed for item-specific memory (target hits vs. unrelated lure false alarms) and gist memory (related lure "hits", i.e., false alarms, vs. unrelated lure false alarms), as proposed by Koutstaal and Schacter (1997). As shown in Table 2.1B, both adjusted measures of true recognition were significantly higher at STM than LTM, while both adjusted measures of false recognition indicate that susceptibility to gist did not change significantly from short-term to long-term testing.

Phenomenological experience

Confidence levels were predictably higher at STM than LTM, but critically, confidence in false alarms to related lures was equivalent across delay ($t(18) =$

0.01, $p = .99$, $d < .01$; Figure 2.2), suggesting that the illusory experience of false recognition widely documented in LTM may in fact be established within seconds of encoding. At STM, confidence in correct responses to all probe types was higher than confidence in all error types. At LTM, however, confidence in related lure false alarms approached confidence for target hits ($t(23) = -1.39$, $p = .18$, d =..20), and actually exceeded the level of confidence in correct rejection of related lures, as shown by a significant response (*yes*, *no*) × delay (STM, LTM) crossover interaction ($F(1,18) = 30.24$, $\eta_p^2 = .63$).

Although confidence in false recognition of related lures did not change over time, the calibration of subjective ratings may have differed when rendered in the context of other recognition judgments at STM and LTM. Experiment 2 thus employed Remember/Know judgments as an alternative way to assess false recognition phenomenology (Lampinen, et al., 1998). Remember/Know judgments specify different states of subjective awareness during memory retrieval (Gardiner, 1988; Tulving, 1985): either remembering vivid, specific details from the experience of a previous event (*remember*) or “just knowing” with certainty that an event occurred without access to any particular details (*know*).

In the DRM paradigm, high rates of *remember* responses to related lures attest to the robustness, and apparent realism, of false memories (Payne, Elie, Blackwell, & Neuschatz, 1996; Roediger & McDermott, 1995; Yonelinas, 2002). Although the rate of *remember* judgments for lures can be selectively attenuated with warnings or modified instructions (Anastasi, et al., 2000; Geraci & McCabe, 2006; McDermott & Roediger, 1998; Neuschatz, Payne, Lampinen, & Toglia, 2001), a sizable measure of “illusory recollection” persists. Evidence in line with dual-process theories of memory suggests that the Remember/Know distinction and self-reports of confidence represent independent constructs (Gardiner & Java, 1991; Rajaram, 1993). Therefore, using Remember/Know methodology to demonstrate similar phenomenology of short-term and long-term false memories would confirm and extend the findings from Experiment 1, implicating a common processing basis for these forms of illusory recognition.

Experiment 2

Method

Participants

32 individuals (18-27 yrs old) participated for course credit or payment. Two participants were excluded for recognition accuracy scores > 2.5 standard deviations from the mean. Three others were excluded for post-experiment questionnaire responses indicating they failed to understand the Remember/Know distinction.

Design and procedure

The method was the same as Experiment 1 except that remember/know/guess judgments replaced confidence ratings. Following each *yes* response to a probe word, participants used their right hand to indicate whether they *remember* the probe word was in the memory set (recollecting something distinctive about studying the word), they *know* the probe word was present (recognizing the word without retrieving specific details of its study), or their response had been a *guess*. Detailed instructions explaining the Remember/Know distinction were adapted from Rajaram (1993). To equate the number of responses on each trial, a display of three boxes appeared following each *no* response to a probe word, prompting an arbitrary right-handed response.

Trimming (see Experiment 1) retained 92% of STM trials and 98% of LTM trials and did not change the pattern of results.

Results and Discussion

<u>Accuracy</u>

Mean math task accuracy during the STM trial retention interval was 0.87. As in Experiment 1, participants made significantly more false alarms to unstudied, related lure probes than to unstudied, unrelated lure probes, at STM ($t(26) = 5.55$, $d = 1.26$) and LTM ($t(26) = 5.87$, $d = 1.13$; Table 2.1A). Again the lure type (related, unrelated) × delay (STM, LTM) interaction was not significant, showing consistency in the false memory effect from short-term to long-term testing ($F(1,26) = 1.55$, $p = .22$, $\eta_p^2 = .06$). Notably, this interaction remains unreliable even when the error data from Experiments 1 and 2 are combined ($F(1,50) = 2.59$, $p = 0.11$, $\eta_p^2 = .05$). The relative time-invariance in susceptibility to gist was corroborated by both adjusted measures of overall false recognition for Experiment 2 (Table 2.1B).

Phenomenological experience

An overall ANOVA on the proportion of *remember* responses (out of all responses) revealed a main effect of probe type ($F(2,52) = 98.14$, $\eta_p^2 = .79$), reflecting higher rates of “remembering” target probes than related or unrelated lure probes. We also found a main effect of delay ($F(1,26) = 12.69$, $\eta_p^2 = .33$), indicating more *remember* responses were made at STM than LTM, and a significant lure type × delay interaction ($F(2,52) = 29.63$, $\eta_p^2 = .53$). As shown in Table 2.1A, the proportion of *remember* responses is significantly higher for related than unrelated lures at both STM ($t(26) = 2.95$, $d = .85$) and LTM ($t(26) = 4.62$, $d = .98$). This difference shows that the overall false memory effect observed in *yes* responses (related lure false alarms - unrelated lure false alarms) is not due solely to *know* responses for which verbatim memory is absent.

The rate of *remember* responses to related lures actually increased at later delays, although this shift was confounded with a rising baseline (unrelated lure) false alarm rate. Normalized estimates of false recollection, expressed as the proportion of *remember* responses to related lures out of the total proportion of *yes* responses to related lures, did not significantly differ from short-term to long-term testing ($t(22) = 0.99$, $p = .33$, $d = .21$; Table 2.2). In other words, the quality of the memory illusion appears equally robust at STM and LTM, indicating that the subjective feeling of certainty which characterizes some instances of false recognition may be relatively time-invariant, consistent with the stable confidence in related lure false alarms in Experiment 1.

As shown in Table 2.2, while the normalized incidence of *remember* responses to related lures is relatively stable over time, it is lower than that for *remember* responses to target probes at both STM ($t(22) = -2.80$, $d = .65$) and LTM ($t(26) = -2.65$, $d = .65$). Unlike the confidence judgments in *yes* responses from Experiment 1, which did not differentiate between related lures and targets at LTM, here a significant difference persists for *remember* responses to these two probe types, suggesting that participants could still use the presence or absence of verbatim detail to distinguish true from false memories at LTM.

In contrast to the selectivity of *remember* responses based in recollective experience, the normalized incidence of *know* responses was equivalent for related lures and target probes at STM ($t(22) = -1.27$, $p = .22$, $d = .27$), and LTM ($t(26) = 1.42$, $p = .17$, $d = .33$). Thus, while *remember* responses comprised a smaller proportion of false recognition than true recognition at both STM and LTM, *know* responses suggest the contribution of familiarity did not differentiate between related lures and targets at either delay. These results indicate that gist memory exerts equivalent influences in STM and at LTM when context information is inaccessible.

Experiment 2 replicated key similarities between memory distortions under STM and LTM conditions, and further characterized the phenomenology of illusory recognition. As in Experiment 1, the corrected false memory rate was stable across delay. Critically, despite higher absolute numbers of false memory errors at LTM, the proportion of false memories associated with “remember” phenomenology was statistically equivalent in the short and long term.

Nevertheless, while LTM confidence ratings in Experiment 1 were equivalent for falsely recognized related lures and correctly recognized target probes, differing proportions of LTM *remember* responses in Experiment 2 imply subtle qualitative differences in the content of true and false memories that persist over time.

General Discussion

Our results extend recent findings of rapid semantic distortions (Atkins & Reuter-Lorenz, 2008; Coane, et al., 2007) and directly connect their occurrence to the well-established phenomenon of false long-term memories (Brainerd & Reyna, 2005; Gallo, 2006). In an STM task, we observed illusory effects that characterize LTM errors: Unstudied theme words are falsely recognized, “remembered” as studied, and endorsed with considerable confidence.

Critically, the normalized incidence of these errors did not change from short-term to long-term testing. The expectation that illusory recognition of semantic associates would be more robust in LTM than STM, based on differential effects of interference, monitoring processes, semantic codes, and access to verbatim traces, was not supported. Instead, false memory rates and phenomenological measurements were found to be relatively stable, suggesting that processes responsible for semantic memory errors are not unique to LTM conditions. These commonalities favor the operation of unitary memory processes across delay.

One reasonable interpretation of our findings that is consistent with memory-based (versus decision-based; see Gallo, 2006) accounts is that false memory errors stem from heightened familiarity of the related lure due to semantic activation from the memory set. Indeed, we found equivalence in the normalized incidence of *know* responses (associated with feelings of familiarity that lack verbatim detail) to related lures and to target probes at both short and long delays. As documented in LTM, the vivid yet illusory qualities of false *remember* responses may result from the process of "content borrowing" (Lampinen, Meier, Arnal, & Leding, 2005), whereby veridical details from studied items are attributed to related but unstudied items. If so, the present results would constitute a demonstration of this effect at the lower temporal boundary of a recognition task (cf. Lampinen, Ryals, & Smith, 2008). Moreover, the present results indicate that illusory recognition can be independent of the declining availability of verbatim traces: the level of confidence and the proportion of *remember* judgments assigned to related lure false alarms are equivalent at STM and LTM.

The immediate vulnerability of memory to distortion indicates the rapid activation of semantic associations (Roediger, et al., 1998; Underwood, 1965) along with an immediate breakdown of monitoring processes, thereby minimizing the potential contributions of decay or other time-dependent processes as sources for these illusory phenomena. Semantic priming may contribute to the effects we observed, as suggested by previous investigations of false memories in implicit memory tests (Cotel, Gallo, & Seamon, 2008; McDermott, 1997; Tse & Neely, 2007); however, nonconscious processes alone offer an insufficient explanation of the phenomenological qualities of false recognition. Furthermore, if conscious generation of the related lure word were occurring at encoding and producing memory traces which could then support false *remember* or high-confidence responses, we might expect a greater proportion of illusory recognition at STM than LTM. However, our results were not in this direction. An alternative possibility is that both associative activation and monitoring operations may be strong under STM conditions, yet counteract each other so that the false memory effect is the same as under LTM conditions when different processes are engaged (Gallo, 2004). Careful investigations varying activation and monitoring levels independently would be required to test this hypothesis. However, the stability of false memory phenomenology across delay supports a semantic rather than associative explanation, because the vividness of a related lure brought to mind by associative processes during encoding, like that of an actually studied item, would be expected to quickly decline (Brainerd, et al., 2008).

Theoretical accounts of false memories have been largely confined to LTM processes, because even when testing is immediate, the length of studied lists typically exceeds working memory span. Likewise, STM is rarely mentioned in theoretical discussions of false memories, suggesting that generally, STM processes are not considered relevant to these effects (however, see Kimball, Smith, & Kahana, 2007). The present results suggest otherwise by demonstrating that false memory errors and associated phenomenology can arise within seconds, indicating that explanations of false memory need not be confined to episodic long-term remembering and should be extended to include short-term effects as well. Our findings are thus consistent with the interpretation that false memories occurring over varying delays may emerge as a consequence of semantic processes that can operate under STM and LTM conditions, and are compatible with unitary models of memory.

Acknowledgments

This chapter represents a collaboration with Patricia A. Reuter-Lorenz, and originally appeared as:

> Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2010). False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1331-1338.

Author Note

This research was supported by the National Institutes of Health (R01AG18286), the National Institute of Mental Health (F31MH079536) and an APF/COGDOP Graduate Research Scholarship.

We are grateful to John Jonides and Cindy Lustig for helpful discussions and advice, and we thank Halle Zucker, Megan Walsh, Laura Zahodne, and Fred Upton for their assistance with stimulus preparation and data collection. Thoughtful comments from Reed Hunt, Charles Brainerd, and an anonymous reviewer improved an earlier version of this manuscript. We also thank the International Max Planck Research School "The Life Course: Evolutionary and Ontogenetic Dynamics" (LIFE) for valuable feedback on this work. Portions of this article were presented at the Annual Meeting of the Cognitive Neuroscience Society, April 2008, in San Francisco.

Figure 2.1 Experiment 1 design. Each four-word list is probed only once; either immediately following a 3-4 second filled retention interval (short-term memory), or in a surprise recognition test after all lists are encoded (long-term memory).

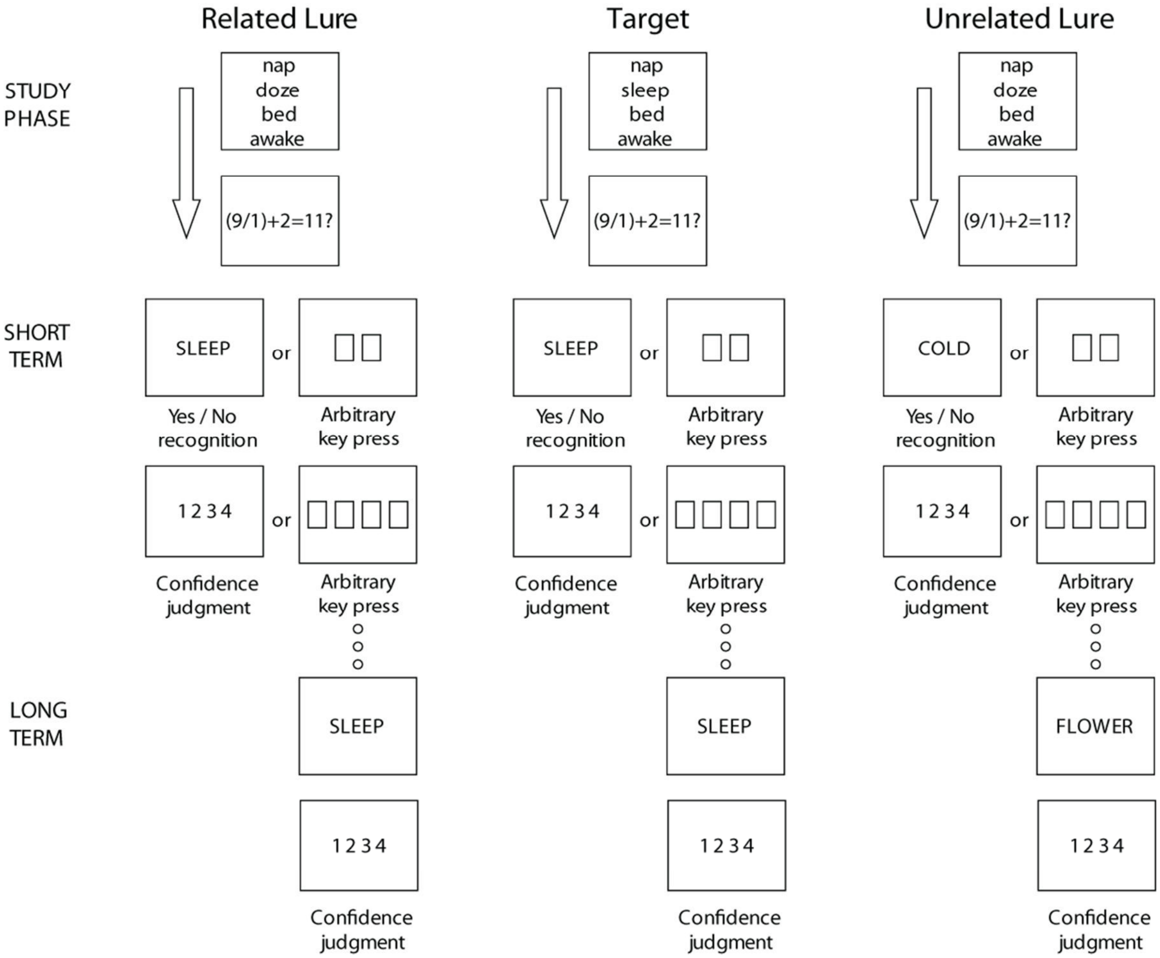


Figure 2.2 Experiment 1 mean confidence ratings by probe type and delay (error bars = SEM; C = correct response). Note that confidence in related lure false alarms (“Yes” responses) is equivalent in short-term memory trials and long-term memory trials.

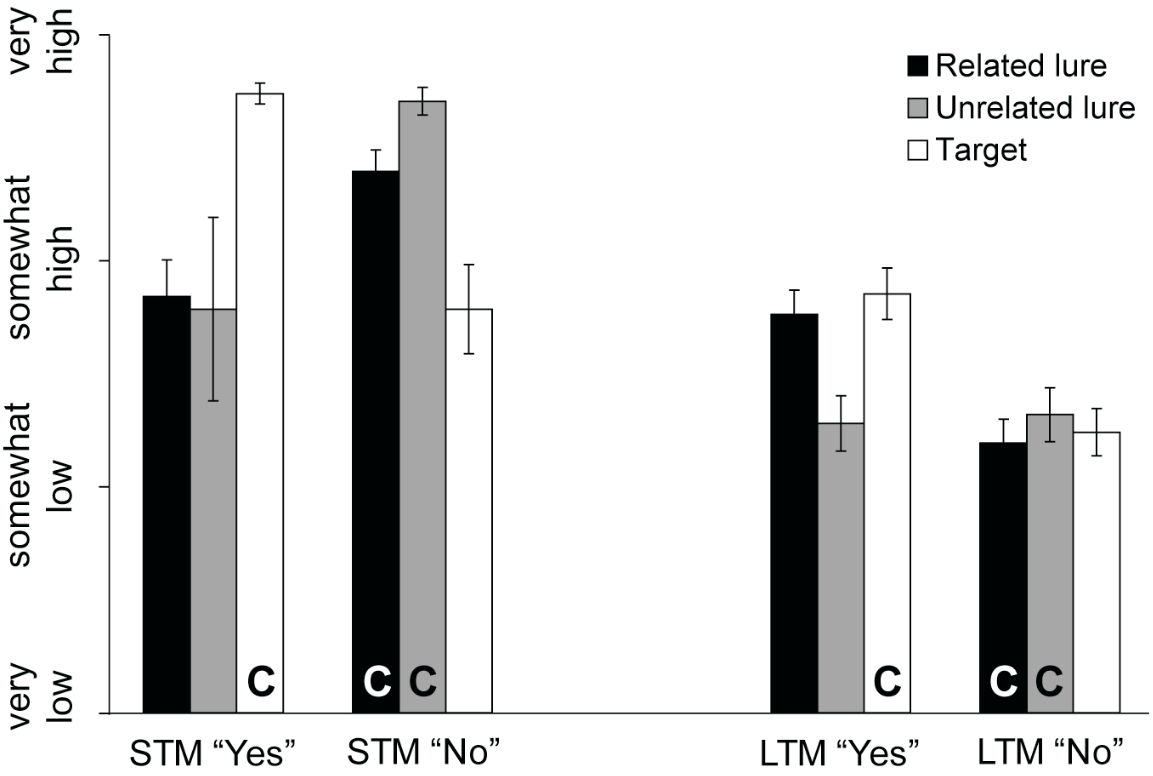


Table 2.1 Recognition performance in Experiments 1 and 2.

A. Mean recognition proportions

	Experiment 1		Experiment 2			
	STM	LTM	STM		LTM	
Probe type	"Yes"	"Yes"	"Yes"	(R/K/G)	"Yes"	(R/K/G)
Related lure	.18	.57	.17	(.05/.05/.06)	.46	(.12/.18/.15)
Unrelated lure	.01	.35	.04	(.01/.01/.02)	.28	(.03/.13/.11)
Target	.92	.69	.88	(.51/.32/.03)	.57	(.23/.19/.12)

B. Measures of recognition performance adjusted for a shifting baseline (unrelated lure) false alarm rate

	Experiment 1					Experiment 2				
	STM		LTM			STM		LTM		
	M	*SEM*	M	*SEM*	*p*-value	M	*SEM*	M	*SEM*	*p*-value
True recognition										
Discriminability (Pr)	.90	*.02*	.34	*.04*	< .001	.84	*.02*	.29	*.04*	< .001
Item-specific memory sensitivity (d')	3.10	*.11*	.91	*.11*	< .001	2.71	*.10*	.78	*.13*	< .001
False recognition										
Discriminability (Pr)	.17	*.04*	.23	*.04*	.31	.13	*.02*	.18	*.03*	.22
Gist memory sensitivity (d')	.78	*.14*	.62	*.11*	.39	.61	*.10*	.52	*.09*	.49

Note: The signal detection measure d' was used to estimate recognition sensitivity, following the recommendation of Seamon et al. (2002), who compared sensitivity measures in a DRM experiment and argued that d' is superior to the nonparametric measure A' for discriminating change in true recognition versus false recognition over time. The pattern of results reported above does not change if A' is used.

Table 2.2 Experiment 2 proportion of "Remember", "Know", and "Guess" responses out of total proportion "Yes" responses.

	STM						LTM					
	"Remember"		"Know"		"Guess"		"Remember"		"Know"		"Guess"	
Probe type	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	.36	*.09*	.29	*.07*	.35	*.08*	.27	*.05*	.39	*.06*	.34	*.06*
Unrelated lure	.32	*.14*	.23	*.12*	.45	*.16*	.13	*.05*	.51	*.07*	.36	*.06*
Target	.60	*.06*	.37	*.06*	.03	*.01*	.43	*.04*	.31	*.04*	.26	*.04*

CHAPTER 3

YOU CAN GO YOUR OWN WAY: EFFECTIVENESS OF PARTICIPANT-DRIVEN VERSUS EXPERIMENTER-DRIVEN PROCESSING STRATEGIES IN MEMORY TRAINING AND TRANSFER

Introduction

Healthy aging brings with it many psychological benefits, which include extending personal growth, wisdom, and generativity, but these features are frequently overshadowed by the stereotype of late adulthood as a time of diminishing cognitive abilities. In the field of cognitive psychology, conclusions about the inevitability of age-related memory decline have been fueled by evidence that older adults often fail to spontaneously engage deep, associative encoding processes that facilitate later memory. However, research has consistently shown that providing instructions and environmental support to encourage such processing can change older adults' memory performance to more closely resemble that of young adults (e.g., Braver, Gray, & Burgess, 2007; Craik & Byrd, 1982; Logan, et al., 2002; Naveh-Benjamin, et al., 2007; Paxton, et al., 2006). These findings are important in indicating that effortful cognitive abilities are not entirely lost with advancing age, but are frequently latent in the absence of external support. Furthermore, they hint at individual differences in strategy use under unsupervised learning conditions, illustrating the wide variety of predispositions and preferences that older adults bring to bear on memory tasks. This is an important consideration for the development of training interventions to maintain or enhance cognitive fitness late into the lifespan, as evidence suggests that older adults' self-generated memory strategies can be at least as effective as experimenter-provided mnemonics (Baltes, et al., 1989; Derwinger, et al., 2003; Hill, et al., 1990) and may produce long-lasting benefits which are less dependent on environmental support (Derwinger, et al., 2005).

The present study tests the hypothesis that encouraging older adults to spend sufficient time and effort encoding information for better memorability will improve memory performance, especially for individuals who might otherwise fail to self-initiate effective encoding strategies. A critical concern in designing effective cognitive training programs for older adults is that individuals with lower initial ability often show the smallest training gains in conventional interventions, despite being the very population most in need of support (Verhaeghen & Marcoen, 1996; Verhaeghen, Marcoen, & Goossens, 1992; Yesavage, Sheikh, Friedman, & Tanke, 1990). While many existing interventions emphasize bolstering the cognitive processes that weaken with age, the possibility of capitalizing on the processes that remain intact – but inefficiently used – has often been overlooked (Park, Gutchess, Meade, & Stine-Morrow, 2007). Explicit strategies or mnemonics provided by an experimenter may aid participants who otherwise would not have thought to employ them, but this one-size-fits-all approach is insensitive to individual differences in existing cognitive strengths and weaknesses, which is a particular concern for low-ability individuals (Calero & Navarro, 2007; Hill, et al., 1989; Verhaeghen & Marcoen, 1996). For example, training with verbal encoding strategies may be suboptimal for individuals who do not already possess strong verbal skills (Yesavage, et al., 1988). This problem is compounded by the finding that while strategy-based training can improve older adults' performance on a target task (e.g., Ball, et al., 2002; Rebok & Balcerak, 1989), the benefit frequently fails to transfer to untrained tasks (see Lustig, et al., 2009, for a review and further discussion). However, sufficient external support in the absence of specific encoding instructions might foster the self-generation of effective memory strategies tailored to participants' individual preferences and pre-existing strengths, and this approach may be more likely to promote transfer to real-world situations.

Older adults' failure to self-initiate effective memory strategies may reflect a *production deficit*, in which strategies are accessible but not spontaneously produced (Dunlosky & Hertzog, 1998; Kausler, 1994; Verhaeghen & Marcoen, 1994), or a *utilization deficit*, in which strategies are produced but not successfully employed (Dunlosky & Hertzog, 1998; Dunlosky, Hertzog, & Powell-Moman, 2005; Jones et al., 2006) – both potentially in addition to, or interacting with, age-related declines in processing capacity (Jones, et al., 2006; Salthouse, 1996). Strategy-based explanations are consistent with findings that older adults can and do benefit from experimenter-prescribed instructions in the context of memory training (Rebok, Carlson, & Langbaurn, 2007; Verhaeghen, et al., 1992). For example, early studies showed that training with mnemonics ranging from method-of-loci (e.g., Rebok & Balcerak, 1989) to pegword (Wood & Pratt, 1987) to face-name association (Yesavage & Rose, 1984) significantly improved older adults' memory performance. Without explicit guidance or external support, however, older adults are often less likely than young adults to self-report using effective memory strategies spontaneously (Dunlosky & Hertzog, 2001; Touron & Hertzog, 2004). And even with instruction, older adults may be less likely than young adults to apply the experimenter-mandated strategy correctly (Verhaeghen & Marcoen, 1996). In summary, evidence suggests that specific mnemonics provided by an experimenter can help some older adults overcome a production deficit, but at the risk of creating a utilization deficit for those with lower initial ability. An understudied question in cognitive training research is whether environmental support itself (such as conditions that encourage sufficient time and effort at encoding) can prompt older adults to engage individualized strategies that will benefit memory performance.

General slowing theories of cognitive aging (e.g., Cerella, 1985; Salthouse, 1996) might predict that age differences in memory performance could be reduced by allowing older adults extra time to produce and utilize effective strategies. For example, Thompson & Kliegl (1991) found that recall accuracy could be equated between age groups when encoding times were on the order of three times longer for older adults than young adults. However, research has also demonstrated that the length of time available in intentional encoding tasks (i.e., presentation duration) interacts with initial ability in predicting older adults' self-generated strategy use. Craik and Rabinowitz (1985) presented young and older participants with word lists at the rate of 1.5, 3, and 6 s per word, under instructions to either simply learn the words for a later test, or to use an experimenter-provided semantic encoding strategy. They found that both age groups benefitted from longer study times to the same extent when environmental support was present, either through encoding strategy instructions or recognition test conditions. When no specific strategy was provided and memory was tested with free recall, it was actually the young adults who showed the greatest benefit from slower presentation rates. Similarly, Hertzog, Dunlosky, and Robinson (unpublished manuscript) measured the frequency of spontaneous strategy use in a paired-associate recall task and observed that, for trials when an effective encoding strategy was reported, young adults outperformed older adults under slow-paced but not fast-paced study conditions. Such findings imply that the availability of extended encoding time alone is not sufficient to raise older adults' memory performance to the level of young adults. Might the availability of *unlimited* encoding time be the key to reducing age differences, by allowing older adults to take as much time as they need to implement effective strategies? Evidence would suggest not, as older adults tend to allocate less time in a study phase than young adults, under self-paced conditions (Dunlosky & Connor, 1997; Murphy, Sanders, Gabriesheski, & Schmitt, 1981) – perhaps due to differential use of information from metacognitive monitoring (Dunlosky, Kubat-Silman, & Hertzog, 2003). Comparable effects are revealed by individual differences within, not only between, age groups: in a memory-training study with unconstrained encoding time, Bissig and Lustig (2007) found that older adults who spent proportionally less time on study phase trials than test phase trials (consistent with the use of reactive rather than proactive control; Braver, et al., 2001; Braver, et al., 2007) showed poorer performance on the training task. Thus, simply supplying older adults with as much study time as they choose is not enough to overcome their deficits in self-initiating effortful memory processes, which indicates that additional environmental support may produce the most beneficial outcomes.

On the basis of previous literature identifying encoding processes as an attractive target for training in older adults, and an initial study revealing the extent to which differences at encoding accounted for training gains under open-ended conditions (Bissig & Lustig, 2007), we designed a memory training intervention which imposed generous study times for all participants and manipulated encoding instructions. Based on the repetition-lag procedure developed by Jennings and Jacoby (2003), the present study examined training and transfer effects in conditions that either 1) mandated a deep, associative encoding strategy, 2) attempted to suppress such encoding by mandating rote rehearsal, or 3) encouraged effort towards encoding (by enforcing study times) but allowed participants to choose their own strategies. Our initial hypotheses (see Lustig & Flegal, 2008) were formed around the consequences of enforcing strategies believed to either benefit memory performance (integrative encoding) or to suppress effective encoding processes (rote rehearsal). We anticipated that instructing one group of older adults to use strategies reported by the most successful participants in our earlier training study might help them to overcome difficulties with self-initiation and potentially minimize the influence of pre-existing ability. The condition in which no specific strategy was provided controlled for the amount of encoding time but was otherwise analogous to the unconstrained setting of our earlier study (Bissig & Lustig, 2007), and we initially expected that a number of these participants would likewise struggle to self-initiate effective encoding strategies without explicit guidance.

Although not a focus of our original predictions, individual differences emerged as an important factor in our analyses, and we discovered that environmental support in the form of fixed encoding time benefitted participants in all three experimental conditions (and was sufficient in itself to diminish age and ability differences in training task performance). We also found that pre-existing ability influenced training gains even when encoding instructions were experimenter-controlled, and that a variable at least as important as how "effective" a memory strategy was deemed to be is whether a memory strategy was mandated or self-selected. There were hints of these effects in our preliminary data (Lustig & Flegal, 2008), and they are consistent with previous research associating self-generated strategies with superior memory performance in older adults (Derwinger, et al., 2005; Derwinger, et al., 2003; Hill, et al., 1990). Of interest, a study by Paxton, Barch, Storandt, and Braver (2006) reported that for older adults, extended practice on a task involving attention to contextual cues was as effective as explicit strategy training for promoting a proactive approach to context processing. Similarly, an earlier study by Murphy et al. (1981) tested older adults in a serial recall task and found that a group who were assigned enforced encoding time outperformed a self-paced group who were instructed to maximize their accuracy by taking as much encoding time as necessary, and even surpassed the performance of another self-paced group who were given explicit strategy training to improve their accuracy. Such outcomes are concordant with findings from the present study which suggest that enforcing ample time-on-task may provide a greater benefit for increasing self-initiated processing than using that same amount of time to train with an experimenter-prescribed strategy.

Method

Participants

Ninety healthy older adults (n =30 per group; 65-92 years of age; 61 female) were assigned to the three training conditions, with the groups matched closely in age, education, and gender (see Table 3.1 for demographics). Within each group, participants were stratified by age in 5-year bins (i.e., an approximately equal number of participants age 65-69, 70-74, 75-79, etc.). All participants were screened for medical or psychological conditions that could influence performance, and had Mini Mental State Evaluation scores (MMSE; Folstein, Folstein, & McHugh, 1975) scores above 24 (mean = 28.4). The study was approved by the University of Michigan Institutional Review Board, and written informed consent was obtained from all participants. None of the participants withdrew from the study prior to completing all eight sessions.

Materials and procedure

Scheduling. Each participant completed eight study visits scheduled over the course of three weeks. The first day included informed consent procedures, a health and demographics questionnaire, dementia screening measures (MMSE and Short Blessed Test; Katzman et al., 1983), the Extended Range Vocabulary Test (ERVT; Educational Testing Services, 1976), baseline (pretest) administration of potential transfer tasks, and brief practice with the training task to familiarize participants with the encoding instructions and time constraints. The last day included posttest administration of the transfer tasks and a questionnaire about strategies used in the training task. This questionnaire was administered at the very end of the last study visit.

Training task. The training task was a modified version of the repetition-lag procedure developed by Jennings and Jacoby (2003; see also Jennings, Webster, Kleykamp, & Dagenbach, 2005). The base procedure consists of 28 study-test “cycles”. In each cycle, participants first study 30 words, presented one at a time. The test phase consists of a yes/no recognition test that includes the 30 studied words as well as 30 unstudied words (lures). The unstudied words repeat within the test list (classified as “new” on their first presentation and “repeated” on their second presentation), requiring participants to discriminate items that are familiar because they were encountered in the study phase from those that are familiar because they were previously-presented lures from earlier in the test phase. A feedback screen appeared after each response, indicating accuracy (correct or incorrect) and trial type (studied, new, or repeated).

The level of difficulty for the retrieval test was dynamically adapted to individual performance by increasing the number of items (“lag”) between lure repetitions in the test phase once criterion performance was reached at the current level. At each level, half of the lure repetitions occurred at a short lag (few items in between repetitions) at which the participant could perform well. These trials were included to help to maintain confidence and motivation. The other half occurred at a long lag (more items in between repetitions), and were used to challenge performance. The possible lag-interval combinations were 1&2, 1&3, 2&4, 2&8, 4&12, 4&16, 8&20, 8&24, 12&28, 12&32, 16&36, and 16&40. All participants started the training task at the easiest level, at which half of the lures repeated after only one intervening word and the other half repeated after two intervening words (i.e., lag level 1&2; see Figure 3.1A). The criterion for advancing to the next lag level was set at 96% correct rejections of long-lag repeated lures for levels up to 2 and 8 (i.e., the 4th level) and relaxed to 93% for higher levels. Once a participant reached the maximum level (16&40), she or he continued working at that level for the remaining sessions. Participants completed four study-test cycles on each of the seven days of training, for a total of 28 training cycles.

Study and test words were chosen from the English Lexicon Project (Balota et al., 2007) and had a mean length of 5.76 letters and mean frequency of 20,487 out of 131 million. Length and frequency were balanced across lists and across conditions (studied, unstudied-short-lag, unstudied-long-lag). Each word was presented in large (32-point Arial) black-on-white font in the center of a computer screen. E-prime software (Psychology Software Tools, Pittsburgh, PA) was used for stimulus presentation and response collection (via key press).

The major differences between the training procedure used here (see also Lustig & Flegal, 2008) and the procedure originally designed by Jennings and Jacoby (2003) occurred at the study phase. These differences included the encoding times used in the study phase of the procedure (2 seconds per word in the original study; 14 seconds per word here), and the assignment of participants to one of three encoding conditions (Integrated Sentences, Strategy Choice, Enforced Rehearsal). The longer encoding time used here was chosen on the basis of the encoding times used by good performers in a previous experiment (Bissig & Lustig, 2007) and on the time needed by a separate, pilot group of older adult participants to implement the Integrated Sentences condition.

In the Integrated Sentences condition, participants were instructed to make a sentence out of each word and (for all but the first word in the list) the word that had just preceded it. Participants were to say the sentences aloud, and they were recorded to ensure compliance and for later content analysis. In the Strategy Choice condition, participants were instructed to think about the meaning of each word presented during encoding in any way they would like, but no explicit strategy was specified. In the Enforced Rehearsal condition, participants were instructed to repeat each word out loud at a fixed pace of once every 2 seconds, guided by a counter on the computer screen. A preliminary report of data from a subset (n =16 per group) of participants in the Integrated Sentences and Strategy Choice conditions appeared in Lustig and Flegal (2008).

The present results generally corroborate that early report and add new findings; any discrepancies are highlighted in the Results section below.

Transfer tasks. A battery of transfer tasks was administered before the first day of training and immediately following the last session. These tasks were used to test the transfer of training benefits to untrained tasks, to help identify which processes were being trained by the intervention, and to assess whether the trained processes and the effectiveness of training differed across encoding conditions. Across all transfer tasks, items were carefully screened to avoid overlap with the training lists and with the stimuli used in other transfer tasks. Alternate forms were used at pre versus post-test for all laboratory cognitive measures except for the Trail-Making Test, for which they do not exist (see below).

Some of the tasks were hypothesized to show transfer effects because they emphasized semantically based and/or integrative processes that were expected to improve if the training task primarily increased deep, associative processing at encoding. One such measure was a shopping-list memory task, in which participants studied a list of 15 individually presented words representing typical grocery items (e.g., “potatoes,” “soup”) and were then given a self-paced old/new recognition test for the 15 studied items and 30 unstudied lure items. On Day 8, half of the “unstudied” items were words that had previously been studied items on the shopping list on Day 1, and half were completely new. On posttest administration, participants were instructed to identify as “old” only those words that had been on the list that day, and to call all other words “new” even if some might have been seen on an earlier day. This manipulation was designed to assess participants' ability to resist proactive interference from pretest to posttest, when previously studied items would be (potentially) familiar but no longer relevant. Another measure on which we predicted positive transfer was a face-name association task, in which participants studied 10 face-name pairs (using digitized black and white photographs, with self-paced encoding and instructions to make a sentence on each trial that connected the name with the face). Participants were then given a two-stage memory test, in which they were first presented with each face and asked to recall the name that had been paired with it; if they could not recall the name, they were given the correct name and a lure name and asked to identify the correct one.

Other tasks were used as "negative controls" to test the hypothesis that general practice or engagement and stimulation as a result of participation in the training program might lead to performance improvements without regard to which processes were targeted for training. These tasks did not emphasize the semantically based or integrative processes targeted by our training procedure, and so were predicted to show little or no improvements. These measures included the Pattern Comparison Test (Salthouse & Babcock, 1991), a common measure of cognitive speed; the Trail-Making Test (Armitage, 1946), which measures cognitive speed (Version A) and executive function (Version B); and pattern and word versions of a self-ordered pointing test (SOPT; Attneave & Arnoult, 1956), a measure of working memory. Our versions of the SOPT consisted of 16 words or 16 patterns arranged in a 4 × 4 grid. There were 16 pages for each test, and the 16 items for that test were arranged differently on each page. On each page, the participant’s task was to point to an item to which he or she had not previously pointed.

Another purpose of the transfer task battery was to assess the potential impact of our training procedure on real-world memory. In addition to measures such as the shopping-list memory task and face-name association task, which were designed to simulate the type of memory tasks that older adults encounter in daily life outside the laboratory, participants were also asked to complete the 35-item Everyday Memory Questionnaire (EMQ; Sunderland, Harris, & Baddeley, 1983) on the first visit and prior to each of their following visits. Participants were asked to indicate how many times within the last 24 hours they had committed each of the memory errors listed on the questionnaire. The EMQ has five subscales: Speech (e.g., “Finding that a word is on the tip of your tongue”), Reading/Writing (e.g., “Forgetting what the sentence you have just read was about and having to reread it”), Faces/ Places (e.g., “Failing to recognize television characters or other famous people by sight.”), Actions (e.g., “Discovering that you have done some routine thing twice by mistake.”), and New Things (e.g., “Forgetting to keep an appointment.”). Participants also completed the Memory Self-Efficacy Questionnaire (MSEQ-4; Berry, West, & Dennehey, 1989) at pretest and posttest. Participants were asked to rate their confidence in performing different memory tasks (e.g., remembering parts of a story or items on a shopping list) at different levels of difficulty (two items, eight items, and so forth). Any inconsistencies observed between EMQ and MSEQ responses could help to rule out placebo effects or factors not directly related to the intervention as explanations for self-reported improvements in everyday memory, because fewer EMQ errors should be correlated with greater confidence in memory ability if the effect were driven by expectancy of post-training improvement.

Three additional transfer tasks were administered only at the end of training (i.e., no pretest) because of the likelihood that knowledge of the test format would influence the strategies participants engaged at encoding. Immediately following the final study-test cycle on the last day of training, participants were given a surprise recognition test for unstudied lures from the final training test list (see Figure 3.1B). This lure test included 15 studied words that appeared on the preceding test, 15 words that had served as lures on the preceding test (i.e., “previous lures”), and 30 completely new words. Participants were now instructed to identify as “old” any word that had appeared in the final test session, regardless of whether it was originally a studied word or if it had been a “new” word (a previous lure item) to be rejected as unstudied at the time. Next, in a false memory test based on the DRM paradigm (Deese, 1959; Roediger & McDermott, 1995), participants viewed three consecutive lists of 12 semantically related words (e.g., *bed*, *rest*, *awake*) in a self-paced study phase and were then given a self-paced old/new recognition test including studied words (e.g., *bed*), unstudied, unrelated words (e.g., *flower*), and unstudied but related “critical lure” words (e.g., *sleep*). The test contained nine studied words (three from each theme list: SLEEP, WINDOW, and SMELL), the three “critical lure” theme words for the studied lists, and 12 unstudied, unrelated words: three items and the critical lure from each of three unpresented lists. The difference between a corrected measure of true recognition (hits to studied items - false alarms to unstudied list items) and a corrected measure of false recognition (false alarms to critical lures - false alarms to unstudied list lures) was used to quantify participants' susceptibility or resistance to false memory (Balota, et al., 1999). Finally, in a word- and source-memory task, participants listened to 30 auditorily presented words, half in a male voice and half in a female voice (counterbalanced across subjects), pseudorandomly intermixed. The time to advance to the next study trial was self-paced, but the auditory stimuli themselves were relatively brief in duration (as they were single words spoken at a normal speed) and could not be repeated. This study phase was followed by a visually presented 60-item recognition test in which participants indicated whether words were old or new and, if old, whether they had been spoken in a male or female voice.

Results

We conducted between-subjects ANOVAs to evaluate the effects of experimenter-prescribed condition and self-reported encoding strategy (see "Encoding condition adherence", below), and used mixed design ANOVAs with repeated measures when assessing training effects and pre- to post-test changes in transfer task performance. Degrees of freedom reported below vary for some of the transfer tasks which were not completed by all 90 participants. The Greenhouse-Geisser sphericity correction was applied as needed to the statistical values reported below, with degrees of freedom rounded to integer values for ease of reading.

Demographics and training-task performance

Participants in the three encoding conditions did not differ in verbal ability (ERVT), speed (Pattern Comparison Test), years of education, or dementia scales (Table 3.1; all *p*s > .17 for main effects of condition [Integrated Sentences, Strategy Choice, Enforced Rehearsal]). They also did not differ in the maximum lag level achieved by the end of the training sessions, *F* < 1.

As previously described (Bissig & Lustig, 2007; Lustig & Flegal, 2008), we ranked training-task performance (1 = *best*, 30 = *worst*) within each condition according to the highest lag level at which participants reached criterion performance, with ties between participants who reached the same maximum lag level resolved by granting the better (lower) rank to whichever participant reached criterion in the earlier training session. Remaining ties were broken by assigning the better rank to whichever participant had the higher overall accuracy for correct rejection of repeated lures. For example, a participant who reached criterion performance for the maximum lag level in Session 15 would receive a better rank than two other participants who both reached it in Session 21, and between those two participants, a better rank would be assigned to the one for whom overall repeated lure accuracy was 90% than the one for whom it was 84%.

In contrast to our earlier study in which encoding was self-paced and unconstrained (“Open-Ended” condition; Bissig & Lustig, 2007), enforcing encoding time diminished the association between poorer training-task performance (indexed by the ranking variable) and older age for all three of the encoding conditions used here (all *r*s < .24, all *p*s > .22)[2]. Performance in the Integrated Sentences condition was significantly correlated with vocabulary and years of education, whereas this was not the case for the Strategy Choice condition; the Enforced Rehearsal condition showed a weaker and nonsignificant relation in this direction (Figure 3.2, Rows 1 and 2). This suggests that participants who had strong verbal skills to begin with were best able to benefit from the compulsory verbal strategy used in the Integrated Sentences condition. In contrast (when paired with the environmental support of long encoding times), the flexibility of the Strategy Choice instructions and the potential for the Enforced Rehearsal instructions to be augmented with other, covert strategies (see below) appears to have allowed participants in those conditions to engage encoding processes tailored to their personal strengths and preferences, thus diminishing the predictive power of pre-existing ability.

Encoding condition adherence*.* We were initially surprised to find that the Enforced Rehearsal group performed as well on the training task as did the other two groups. This condition was originally designed to serve as a control condition that would suppress deep encoding processes and mimic the strategies reported by the least successful participants in our previous study (Bissig & Lustig, 2007). Examination of the post-training strategy questionnaire revealed that many participants in the Enforced Rehearsal condition were covertly engaging deep encoding strategies while also completing the overt rehearsal task (e.g., “I made

---

[2] The age range and mean age of participants in the present study (65-92 years of age, *M* = 75.5, *SD* = 6.8) did not significantly differ from the earlier study (67-93 years, *M* = 74.5, *SD* = 6.1).

associations to the words wherever possible"; "I related each word to something or someone"; "I tried to make a story out of the words").

To better understand the effects of participant-chosen encoding processes, for all three encoding conditions, responses to the question "Did you use any strategy to help you learn the words during the study part of each session?" on the post-training questionnaire were coded as a) shallow (focusing on surface features of the words, rehearsal only, or vague descriptions such as "concentration" or "used good judgment"), b) intermediate (using mental imagery), or c) deep (incorporating the words into a sentence or story, self-referential processing, creating associations to other words, or using other semantically based strategies). Coding was done by two independent raters and any discrepancies were resolved by discussion and arriving at consensus. If a participant reported multiple strategies, the one with the "deepest" level of encoding was counted.

Collapsing across condition, deep encoding strategies were reported most frequently (n = 65), followed by shallow encoding strategies (n = 17). Strategies that fell at an intermediate level of encoding were rare (n = 8), and the performance of the participants who reported them tended to lie between those in the other two groups. For clarity of interpretation, our analyses below exclude these few participants and focus only on the comparison of participants who reported deep or shallow encoding[3].

---

[3] These groups did not differ on demographic variables including age, gender, years of education, or ERVT score (all *p*s > .21). However, participants who reported using shallow encoding had significantly higher (i.e., worse) SBT scores (*M* = 2.24) than participants who reported using deep encoding (*M* = 1.20), *t*(80) = 2.24, *p* < .03.

The maximum lag level achieved by the last day of training did not significantly differ between participants who reported using deep (*M* = 30.1) versus shallow (*M* = 23.3) encoding (*t*(80) = 1.58, *p* = .12). However, a marginally significant day (1,2,3,4,5,6,7) × level of encoding (shallow, deep) interaction showed a trend for deep encoders to reach the maximum possible lag earlier in the training sessions: *F*(2,158) = 2.63, *p* = .08. This pattern was also evident in the number of sessions that participants required to achieve their individual maximum lag levels: deep encoders reached asymptote earlier (*M* = 5th training day) than shallow encoders (*M* = 6th training day); *t*(80) = 2.24, *p* < .03. Thus, training-task progress benefitted from deep encoding strategies, whether they were explicitly instructed (Integrated Sentences), discovered under open-ended conditions with environmental support (Strategy Choice), or covertly implemented in spite of an enforced shallow encoding task (Enforced Rehearsal). Within the self-reported deep encoders, neither the maximum lag level achieved by the last day of training nor the number of sessions required to achieve individual maximum lag levels significantly differed by encoding condition assignment (both *F*s < 1).

Contributions at retrieval

Although the focus of our experimental manipulation was on encoding processes, reaction times (RTs) for correct rejections of lure items during the training task test phase also reveal effects on the recollective processes engaged at retrieval. Decreases in the RT difference between short-lag and long-lag repeated lures would suggest that training helped participants restrict memory access to target information, perhaps as a result of improved encoding processes. In this “source-constrained retrieval” mode (Jacoby, Shimizu, Daniels, & Rhodes, 2005), the lag interval between repetitions should have little effect on correct rejection RTs, as efficiency in rejecting long-lag items is enhanced. A Condition (Integrated Sentences, Strategy Choice, Enforced Rehearsal) × Day (First, Last) × Item Type (new, short-lag, long-lag) ANOVA on the RTs for correctly-rejected items revealed a significant Day × Item Type interaction, $F(2, 174) = 8.02$, $p < .005$, that did not interact further with Condition, $p > .17$. Follow-up analyses showed that while participants were initially slower on new items than short-lag items (1901 ms vs. 1755 ms),by the last day of training, RTs for the two item types were equivalent (1530 ms vs. 1525 ms). Within the repeated item type, however (and replicating our earlier analyses; Lustig & Flegal, 2008), the differential time needed to reject long- versus short-lag items significantly decreased from the first day (2088 ms vs. 1755 ms) to the last day (1612 ms vs. 1525 ms), $F(1, 87) = 33.84$, $p < .0005$. In other words, although response time generally improved from the first to last day of training, the largest increase was in the efficiency to reject long-lag items. Importantly, this did not reflect an increasing bias to reject items as unstudied. Speed to accept studied items also increased from the first to last day of training, (1794 vs. 1540 ms, $t(89) = 6.96$, $p < .0005$), and accuracy remained stable ($M = .85$ at both occasions). In addition, overall better accuracy on studied items correlated with better training rank for all groups (rs > .40, ps < .03), indicating that training-task performance was related to good encoding of the studied items.

Participant-chosen Encoding Level also did not affect the speed at which repeated versus new lures were correctly rejected, *F* < 1. However, when comparing the RTs for correct rejections of short- versus long-lag items, there was a marginal Item Type (short-lag, long-lag) × Day (First, Last) × Encoding Level (shallow, deep) interaction, *F*(1,80) = 3.31, *p* = .07. Increases in the efficiency of rejecting long-lag items across training sessions were somewhat more pronounced for participants who self-reported deep encoding strategies (2180 ms on Day 1 vs. 1666 ms on Day 8, as opposed to 1805 ms on Day 1 vs. 1463 ms on Day 8 for shallow encoders). Again, this increased efficiency did not appear to be due to a generic increase in the tendency to classify items as unstudied; the correlation between better rank and better accuracy on studied items was significant for both shallow and deep encoders (both *p*s < .01).

In addition to the speed of rejecting the different lure types, differences between participants might also arise in how they processed the test phase feedback. In our earlier studies, we found that those participants who spent a proportionally longer time viewing feedback screens that followed incorrect responses than those that followed correct responses ((incorrect feedback RT - correct feedback RT)/(incorrect feedback RT + correct feedback RT) had better training ranks. This effect was replicated here for all three training conditions (Figure 3.2, Row 3), although it was numerically weaker in the Integrated Sentences condition than in the others. When the data were analyzed by participant-chosen Encoding Level, participants who self-reported deep encoding strategies also showed this correlation (*r* = .58, *p* < .001), but it was reduced and not statistically significant for participants who self-reported shallow encoding strategies (*r* = .31, *p* = .23). This pattern is consistent with our interpretation that participants who used shallow encoding were less likely to self-initiate controlled processing, even in response to feedback. Although no significant differences were observed across experimenter-determined training condition, participants who self-reported deep encoding strategies spent proportionally more time viewing feedback after incorrect than correct responses (*M* = 0.35) compared to participants who self-reported shallow encoding strategies (*M* = 0.25), *t*(80) = 2.23, *p* < .03.

Transfer task performance: Changes from pre- to post-test

Shopping list*.* When the data were analyzed by Condition, there was a significant *negative* main effect of pre versus post-test, *F*(1,87) = 10.21, *p* < .001, but no interaction with Condition, *F* < 1. This indicates that overall recognition accuracy was lower on the second (posttest) administration of this transfer task than on the first (pretest). This drop in accuracy was likely due to the interference manipulation introduced on the second administration: In order to assess participants’ ability to reject familiar but currently-irrelevant items, half of the lure items on the second administration were studied items on the first administration (i.e., they had been part of the shopping list on Day 1), and the other half were completely new. However, participants were instructed to identify as “studied” only list items from that day (i.e., Day 8), and to reject all other items (both previously-studied and never-studied lures). A closer examination of the data was consistent with this idea: At posttest, accuracy for items studied at pretest and then used as (familiar) lures was worse than for new, never-studied lures, *t*(89) = 3.42, *p* < .005, and also worse than pretest lure item accuracy (*t*(87) = 4.38, *p* < .001). Accuracy on studied items did not change between the two administrations, *F* < 1, and correct rejection of new lure items decreased, but not significantly, *t*(87) = 1.76, *p* = .08 (Table 3.2). There were no interactions with Condition on any of these measures.

When the data were analyzed by participant-chosen Encoding Level, there was a trend for a Day (pretest, posttest) × Encoding Level (shallow, deep) interaction for overall accuracy, *F*(1, 79) = 3.55, *p* = .06. This interaction was significant for the correct recognition of studied items, *F*(1, 79) = 4.83, *p* = .03, and the correct rejection of new lure items, *F*(1, 79) = 4.66, *p* = .03. In both cases, participants who reported using shallow encoding showed greater declines than those who reported using deep encoding; the latter group showed little or no change. Interestingly, the difference between correct rejection of completely new versus previously-studied lure items trended in the opposite direction, *F*(1, 80) = 3.04, *p* = .09, such that shallow-encoders had generally poor performance that did not differ between the item types, whereas deep-encoders performed worse on previously-studied than on never-studied lures, *t*(64) = 4.21, *p* < .0005.

These declines in performance were unexpected. One possible explanation is that training led participants to use ineffective encoding processes, an obviously undesirable result. Another, more optimistic interpretation is that the declines in performance reflected proactive interference from the pretraining list to the posttraining list, and that the different encoding methods participants adopted during training and during the shopping-list task influenced their vulnerability to this interference. That is, shallow encoders may have suffered a general effect of proactive interference reducing their performance across all measures. In contrast, deep encoding may have helped preserve performance for studied and new items, with the downside of increasing relative vulnerability to semantically-related or previously-studied items (which would have had a strong, difficult-to-reject memory representation due to being deeply encoded themselves). Consistent with this explanation, deep encoders with good training-task performance showed larger proportional differences between previously-studied and never-studied lures at Day 8, $r = .30$, $p < .02$, suggesting that the more successful they were in deploying deep encoding strategies, the more likely they were to incorrectly endorse previously-studied but now-incorrect items on a later test. This correlation was not present for shallow encoders, $r = .01$. As described below, a general pattern emerged across several transfer tasks indicating that deep encoders might show benefits to veridical memory but also increased vulnerability to false recognition of related or previously-studied lures.

Face-name recall. Face-name recall accuracy showed a significant improvement overall, $F(1,87) = 12.33$, $p < .001$, but did not interact with Condition, $F < 1$ (Table 3.2). When the data were analyzed by participant-chosen Encoding Level, deep encoders showed significant improvements from pre- to post-test, $t(64) = 3.57$, $p < .005$, whereas shallow encoders did not, $t < 1$. However, the interaction was not statistically significant, $F(1,80) = 2.49$, $p = .12$.

Correlations between face-name recall improvement and training rank were not significant for either group, both *r*s < .15.

Nonverbal, nonmemory tests. The nonverbal, nonmemory tests used as negative controls for practice effects (Pattern Comparison, Word SOPT, Pattern SOPT) did not show any differences from pre- to post-test or any interactions with Condition, all *F*s < 1 (Table 3.2), consistent with the prediction that little or no improvements would be found on transfer tasks that do not emphasize the semantically based or integrative processes targeted by our training procedure. A general increase in speed from pre- to post-test was found on the Trail-Making Test, for both Version A, *F*(1, 86) = 6.95, *p* < . 05, and Version B, *F*(1, 86) = 12.50, *p* < .005, but neither interacted with Condition, both *p*s > .30. The same patterns were found when these tests were analyzed by Encoding Level.

Notably, the Trail-Making Test was the only nonverbal test for which we did not have alternate forms, suggesting that the improvements seen here were simple practice effects. This may also explain why Jennings et al. (2005) found improvements on SOPT after repetition-lag training where we did not: their experiment used the same forms on both pre- and post-test, whereas we used alternate forms that reduced the influence of general practice or familiarity.

Posttest-only tasks

Surprise lure test. As described earlier (see Methods), after the final study-test cycle on the last day of training, participants were administered a surprise old/new recognition test for the items seen in that cycle. On this test, they were asked to respond “old” to any word that had appeared in the previous training cycle, regardless of whether it had been one of the words on the studied list or one of the unstudied lures from the test phase. This test was intended to be diagnostic of how encoding strategies (experimenter-directed or participant-chosen) might influence retrieval strategies during the training task. In particular, if the Integrated Sentences condition was successful in leading participants to create a strong, integrated representation of the studied list, and thus quickly reject lure items during the test phase, then they should show relatively poor memory for those items as compared to studied items on this surprise test.

Unsurprisingly, participants in all three conditions recognized proportionally more studied items than previous-lure items from the final training session ((studied - lure) / (studied + lure)). As predicted, this difference was greater in the Integrated Sentences condition than in the other two conditions, $F(2, 78) = 3.90$, $p < .05$. The groups did not differ in their correct rejection of new items, or in accuracy for studied items, *F*s < 1 (Table 3.3). No differences were found when performance was analyzed by the participant-chosen Encoding Level, all *p*s > .20. Taken together, these patterns suggest that the processing requirements of the Integrated Sentences condition were generally effective in training participants to build an integrated representation of the studied list and avoid deep processing of lures when they appeared in the test list. Further, although the processing required by the experimenter-provided Integrated Sentences instructions might be characterized as deep or elaborative, this strategy apparently led to different results than did participant-chosen deep encoding methods. In the absence of explicit instructions to perform an integrative encoding task, older adults may be more likely to generate strategies which emphasize item-specific processing more than cross-item connections, a tendency that is consistent with associative-deficit theories of cognitive aging (Naveh-Benjamin, 2000) as well as previous empirical findings (e.g., Luszcz, Roberts, & Mattiske, 1990).

DRM false memory test. In the DRM task, an index of false memory relative to each participant's level of true memory was calculated as the difference between corrected rates of true recognition (hits to studied items - false alarms to unstudied list items) and false recognition (false alarms to critical lures - false alarms to unstudied list lures). Positive values indicate resistance to false memories (true memory > false memory); negative values indicate susceptibility. Mean values did not differ across the experimenter-prescribed conditions or participant-chosen encoding levels, both *p*s > .66 (Table 3.3).

Despite the lack of mean differences between the encoding conditions, we hypothesized that they might differ in their patterns of individual differences. Specifically, the Integrated Sentences condition strongly emphasized making associations between studied items; participants who were especially successful in this kind of relational encoding might be good at recognizing studied items and rejecting unrelated lures, but vulnerable to falsely remembering related (associated) lures (cf. Hege & Dodson, 2004; McCabe, et al., 2004). Exploratory analyses on DRM task performance and training-task rank were consistent with this possibility. All three groups showed significant correlations between training rank and accurate memory (hits to studied items - false alarms to unrelated lures;

*r*s > .37, *p*s < .05). However, while training-task rank was correlated with better resistance to false memory for the Strategy Choice (*r* = -.44, *p* < .02) and Enforced Rehearsal (*r* = -.57, *p* < .001) conditions, this was not the case for the Integrated Sentences condition (*r* = .05, *p* = .82) (Figure 3.3, Row 4). A Fisher *r*-to-*z* transformation analysis confirmed that the magnitude of the Integrated Sentences correlation differs significantly from the Enforced Rehearsal correlation (*z* = 2.56, *p* = .01) and marginally from the Strategy Choice correlation (*z* = 1.92, *p* = .05).

Furthermore, it appears that the lack of an association between greater resistance to false memory and better training task performance was restricted to those participants who followed the experimenter-provided Integrated Sentences strategy (with its emphasis on inter-trial associations), and was not a consequence of deep encoding in general. Collapsed across condition, participants who self-reported using deep encoding showed significant correlations between larger training-task improvements and better resistance to false memory (*r* = .30, *p* < .02). These differences are consistent with the notion that deep encoding need not always be relational in nature (Einstein & Hunt, 1980; Hunt & Einstein, 1981). In cases where relational processing is redundant with the structure of related memory sets, as in DRM lists, deep encoding focused on distinctive characteristics – rather than similarities or associations – may be most beneficial for accurate memory and resistance to false memory (Hunt & McDaniel, 1993), and it is possible that distinctive and/or self-referential processing was preferred by deep encoders in the Strategy Choice and Enforced Rehearsal conditions.

Although Integrated Sentences instructions impaired rejection of related lures in the DRM task, they appear to have supported the rejection of *unrelated* lures in the training-task test phase. Further exploratory analyses showed that the correlation between susceptibility to false memory and poorer recognition of previous lures in the surprise Lure Test that followed the final training session was significant for the Integrated Sentences condition ($r = .49$, $p < .02$) but not for the Strategy Choice ($r = .15$, $p = .44$) or Enforced Rehearsal ($r = .09$, $p = .62$) conditions (here, the magnitude of the Integrated Sentences correlation does not differ significantly from Strategy Choice or Enforced Rehearsal; both *p*s > .10). For Integrated Sentences participants only, the poorer their resistance to false memory in the DRM task, the less likely they were to recognize lures from the final training test list – both predicted outcomes of forming strong associations at study through relational encoding. Together, these findings are consistent with the highly speculative but interesting possibility that an experimenter-enforced strategy to increase deep, associative processing at encoding produces the benefit of decreased lure processing at retrieval, but at the cost of increased susceptibility to false memory when lures are related.

Word and source memory. Because the different training conditions emphasized different aspects of word encoding, differences between them should be expected for word memory but not for source memory (which was not targeted for training in any of the conditions). This was indeed the case: The three groups significantly differed on item recognition (word memory; $F(2, 79) = 3.92$, $p < .03$), with no group differences in source memory, $F < 1$. Post hoc *t* tests showed that word memory accuracy in the Integrated Sentences condition (*M* = 72%) was lower than in the Strategy Choice (*M* = 77%; *p* = .08) or Enforced Rehearsal (*M* = 79%; *p* < .01) conditions (Table 3.3). This difference may have arisen if self-generated strategies (from the Strategy Choice and Enforced Rehearsal conditions) were more likely to be transferred to the word memory task than experimenter-provided strategies (from the Integrated Sentences condition), especially since the per-word presentation times in the word memory task were much shorter than in the training task.

Consistent with this explanation, when the data were analyzed by the participants' self-reported Encoding Level, there was a marginal effect favoring deep encoders for word memory, *F*(1,74) = 3.23, *p* = .08. When participants who were in the Integrated Sentences condition were eliminated from this analysis (because the short presentation times likely discouraged use of the time-consuming sentences strategy), the effect grew stronger, *F*(1,47) = 4.78, *p* < .03. Encoding level did not affect source memory accuracy regardless of whether or not the Integrated Sentences condition was included, all *p*s > .21.

Correlations between training rank and word memory accuracy were high for all conditions (*r*s > .40, *p*s < .04). There were also strong correlations between training rank and source memory accuracy for participants in the Integrated Sentences (*r* = .55, *p* < .01) and Enforced Rehearsal (*r* = .64, *p* < .001) conditions, with a weaker relationship for Strategy Choice participants (*r* = .29, *p* = .14). For the Strategy Choice condition, a significant correlation between rank and word memory, with no correlation between rank and source memory, replicates the pattern observed at n = 16 in our earlier analyses (Lustig & Flegal, 2008). It may reflect a tendency for these participants to transfer the successful strategies they generated during training to a novel word memory task. In contrast, source memory, is not likely to benefit from transferred strategies because it was not a focus of the training intervention.

When the data were split by self-reported Encoding Level, for deep encoders training rank was a strong predictor of both word and source memory (*r* values of .49 and .48, respectively, *p* < .0005). For shallow encoders, training rank was a strong predictor of word accuracy, *r* = .57, *p* < .03, but not source accuracy, *r* = .32, *p* = .22. Consistent with the conclusions drawn from the ANOVA results, dropping Integrated Sentences participants from the analysis slightly increased rank-word accuracy correlations for both deep and shallow encoders (*r* = .54 and *r* = . 67) with little effect on source-memory correlations (*r* = .41 and *r* = .29).

In summary, although the use of a posttest-only design for these measures prevents us from making strong causal statements, these results do help link the training procedure to processes involved in word memory. The experimenter-designated training groups differed in word memory but not source memory, and participant-chosen encoding level also was more strongly related to word memory than source memory. In addition, the different results for participants trained in the Integrated Sentences condition as opposed to the other conditions suggests that training in this very time-dependent manner did not benefit – and may have even impaired – performance in a situation where exposure to the stimuli was time-limited. These results are consistent with the general conclusion that while training may benefit specific cognitive processes (in this case, those related to verbal memory), participant-chosen methods for training those processes may be more effective in producing transfer benefits than are experimenter-chosen strategies.

Self-report measures

Everyday Memory Questionnaire (EMQ). When the data were analyzed by Condition, there was an overall reduction in self-reported everyday memory errors from pre- to post-test, $F(1, 87) = 39.81$, $p < .0005$, but the size of the reduction was smaller for the Integrated Sentences condition than for the other two groups, $F(2,87) = 3.24$, $p < .05$ (Table 3.2). Participant-chosen Encoding Level did not interact significantly with the reduction in everyday memory errors, $F(1, 80) = 1.72$, $p = .19$. Unlike the results for word memory (see above), eliminating the Integrated Sentences participants from the analysis did not change this pattern. These results suggest that allowing participants to choose and practice their own strategies during the training period (even covertly, if task demands imposed by the experimenter-prescribed strategy are relatively low, as in our Enforced Rehearsal condition) was more effective in reducing real-world, everyday memory errors than was prescribing a specific strategy. However, in everyday life (where participants presumably have more control over their environment and learning conditions than in the lab), the exact nature of the self-chosen strategy (i.e., involving a deep or shallow level of encoding) may not have had much influence. In other words, practicing a strategy that the participant was comfortable with and preferred to use appeared to be the important factor for improving everyday memory.

The number of EMQ errors on Day 1 (pretraining) was significantly correlated with training rank only for participants in the Strategy Choice condition (Figure 3.3, Row 1), replicating our previous findings. Also replicating our previous results, training in this condition eliminated the rank-EMQ correlation on Day 8 (Figure 3.3, Row 2), suggesting that those participants who had the most EMQ errors on Day 1 showed the largest benefits from training. The opposite pattern of results was found for the Integrated Sentences condition, where rank did not correlate with EMQ on Day 1, but did on Day 8 (Figure 3.3, Rows 1 & 2). This suggests that in this condition, participants who showed the smallest training gains in the laboratory also showed little benefit in everyday memory by the end of training, perhaps because the experimenter-prescribed strategy was a poor match for their abilities and preferences. Of note, a significant correlation ($r = .65$, $p < .05$) between proportional change on the EMQ and rank for the Strategy Choice condition at n = 16 in our earlier analyses (Lustig & Flegal, 2008) still trends in the same direction but is no longer significant ($r = .27$, $p = .15$) with the complete sample of n = 30 (Figure 3.3, Row 3). Removing one Strategy Choice participant with an outlying EMQ-change score (i.e., p(EMQ change) = -.44 for this individual, compared to $M = .46$, $SD = .35$, for the entire group) restores the significance of the correlation for this condition, $r = .39$, $p < .05$.

When the data were analyzed by participant-chosen Encoding Level, deep encoders showed a small correlation between training rank and EMQ errors on

Day 8 ($r = .29$, $p < .05$, those with worse training ranks reported more errors), but no correlations with Day 1 errors or with the proportional change in errors over training (both *r*s < .15, *p*s > .30). Inspection of the scatterplots suggested that the Day 8 correlations may have been due to two outlying participants. For shallow encoders, there was no relationship between training rank and the Day 1 or Day 8 errors, but a significant correlation between training rank and EMQ change ($r = .51$, $p < .05$). This suggests that participants who failed to self-initiate successful encoding strategies over the course of training were more likely to have entered the study with a large number of reported memory errors on the first administration of the EMQ, with more room for improvement after training

Memory Self-Efficacy Questionnaire (MSEQ). Given that the EMQ is a self-report measure, a potential concern is that improvements on it may reflect a placebo effect. That is, did participants report fewer memory errors because they knew that they were in a memory-training study and thus thought they "should" show a better memory in their everyday lives? The differences in improvement for the three different conditions, described above, argue against this alternative interpretation. To further examine this possibility, we also examined changes in the MSEQ. If fewer reported everyday memory errors on the EMQ reflected a placebo effect, then confidence in memory ability should likewise show an increase, and should be correlated with improvements on the EMQ.

This was not the case. When the MSEQ data were analyzed using a Condition × Pre/Post repeated-measures ANOVA, there was no main effect for either factor, both *p*s > .25. The interaction was marginally significant, $F(2, 87) =$

2.31, *p* = .11. Interestingly, although they were assigned to a condition designed to suppress strategies that would improve memory performance, participants in the Enforced Rehearsal condition showed an increase in MSEQ, *t*(29) = 2.46, *p* < .05, that was not shared by the other two groups, both *t*s < 1. This effect appears to be driven by those Enforced Rehearsal participants who self-reported using deep encoding strategies, *t*(17) = 3.13, *p* < .01, as MSEQ scores did not increase significantly for the other participants in the Enforced Rehearsal condition (*t* < 1). Speculatively, the deep encoders may have experienced a boost in memory self-efficacy from successfully engaging covert, associative encoding processes at the same time as overtly following the less effective experimenter-provided strategy, and discovering a resultant benefit to memory performance.

Importantly, EMQ scores (pretraining, posttraining, and proportional change) did not differ between the Enforced Rehearsal participants who did vs. did not report using deep encoding, arguing against a change in self-efficacy accounting for fewer reported EMQ errors. In addition, when the data were collapsed across experimenter-directed encoding condition, there were no MSEQ differences related to self-chosen Encoding Level, F < 1. Furthermore, there were no correlations between proportional change on the EMQ and on the MSEQ, regardless of whether data were analyzed over all groups, divided by Condition, or divided by participant-chosen Encoding Level (all *r*s < .20, *p*s > .30). Thus, there is no evidence to suggest that changes in perceived memory efficacy that might potentially be caused by a placebo effect of training contributed to the reduction in self-reported memory errors on the EMQ.

Moreover, there were no significant correlations between proportional change on the MSEQ and training-task performance.

Discussion

The present results confirm many of the preliminary conclusions supported by our earlier analyses (Lustig & Flegal, 2008), and extend them with the addition of data from the Enforced Rehearsal encoding condition. In particular, the large size of our complete sample (n = 30 per condition) helped to reveal important differences between the effects of experimenter-directed and participant-chosen encoding strategies.

Relative to a previous experiment in which encoding was self-paced and unconstrained (Bissig & Lustig, 2007), enforced study times in the present study benefitted training-task performance in all three experimenter-determined encoding conditions. Even in the Enforced Rehearsal condition, which was originally designed to serve as a control condition that would suppress deep encoding processes, sufficient study time was evidently available for many participants to covertly engage in deep processing while also completing the overt rehearsal task (as discussed below). As a result, environmental support to facilitate the self-initiation of effortful memory processes diminished the association between older age and poorer training task performance that was present under open-ended conditions (Bissig & Lustig, 2007). However, training-task success was significantly correlated with vocabulary and years of education only for participants in the Integrated Sentences condition, suggesting that the verbal encoding strategy was most beneficial for participants who already possessed strong verbal skills.

Encoding condition assignment also influenced performance on self-report measures and unpracticed transfer tasks. In the Strategy Choice condition, training-task rank was correlated with the number of memory errors reported on the EMQ on Day 1, but not on Day 8 – suggesting that participants in this condition who experienced the most memory errors pre-training experienced the largest benefits from training. In contrast, in the Integrated Sentences condition, training-task rank was correlated with the number of EMQ memory errors on Day 8, but not on Day 1 – suggesting that participants in this condition who showed the smallest training gains in the laboratory also experienced little benefit in the reduction of memory errors posttraining, perhaps because the compulsory verbal strategy was a poor match for their abilities and preferences. Furthermore, training with a deep, associative encoding strategy may have carried the cost of vulnerability to false memory for related but misleading information, in addition to the benefit of higher accurate memory for relevant information. Although training-task rank was significantly correlated with accurate memory in the DRM task for all three encoding conditions, only participants in the Integrated Sentences condition did not show a similar relationship between better training-task performance and greater resistance to DRM false memory. These results are consistent with the hypothesis that more benefits and fewer costs accrue when environmental support is provided but participants are allowed to generate their own encoding strategies.

Self-reported level of encoding, independent of experimenter-prescribed encoding condition, was associated with training-task success. Participants whose post-training questionnaire responses indicated that they used deep encoding strategies advanced through the levels of the repetition-lag task more quickly than participants who reported using shallow encoding strategies. In the DRM task, collapsing across condition assignment, the relationship between better training-task performance and resistance to false memory was significant for participants who reported using deep encoding strategies, suggesting that the lack of this correlation in the Integrated Sentences condition was attributable to transfer of the enforced relational processing strategy, and not a consequence of deep encoding in general. On the posttraining word and source memory task, a differential strategy benefit was also found in that training-task rank predicted source accuracy for deep encoders, but not shallow encoders.

Our findings converge with others from the cognitive aging literature which point to metacognitive monitoring deficits, rather than a lack of knowledge about effective strategies, as a major factor in accounting for older adults’ failures to self-initiate effortful memory processes. The Strategy Choice condition in the present study differed from the Open-Ended condition in a previous experiment (Bissig & Lustig, 2007) only in the environmental support provided by fixed encoding time, but this factor proved to be sufficient to diminish age and ability differences in training task performance that were present when encoding time was self-paced. Deficient monitoring skills may explain why older adults can benefit from extended time, as our results and others have shown (e.g., Murphy, et al., 1981; Paxton, et al., 2006), yet are unlikely to allocate as much time as they need when left to their own devices (Bissig & Lustig, 2007; Dunlosky & Connor, 1997; Murphy, et al., 1981). This interpretation was supported by a study in which monitoring training combined with memory strategy training produced associative memory improvements for older adults, relative to strategy training alone – when the study phase was self-paced, but not experimenter-paced (Dunlosky, et al., 2003). Similarly, an early study by Murphy and colleagues (1981) showed that older adults given enforced encoding time but no strategy training actually performed better in a serial recall task than older adults provided with explicit strategy training, suggesting that an age-related monitoring deficit impedes efficient (and sufficient) use of study time under self-paced conditions.

Hertzog, Price, and Dunlosky (in press) gave older and young adults practice that was either supervised (i.e., experimenter-mandated, using rote repetition and interactive imagery strategies) or unsupervised (i.e., freely-chosen, following exposure to a list of potential strategies) in an associative memory task with fixed encoding time. For a second study period, all participants were free to choose their own encoding strategies. Young adults' recall performance improved (reflecting knowledge updating and a shift toward using more effective strategies) while older adults' did not (indicating a tendency to stick with previous, suboptimal strategies). This result might be considered another example of deficient monitoring skills, with older adults failing to monitor the efficacy of different memory strategies and regulate their behavior accordingly (see also Brigham & Pressley, 1988).

It is worth noting that the memory strategy use reported in Hertzog, Price, and Dunlosky (in press) was freely-chosen, but not strictly spontaneous. The authors acknowledge that pre-exposure to a list of potential strategies in their experiment may have prompted some participants in the unsupervised learning condition to start off with more successful strategies (e.g., interactive imagery) than they would have otherwise thought to use on their own (cf. Dunlosky & Hertzog, 2001). An open question, then, is what determines how and when older adults will spontaneously produce effective memory strategies? Evidence suggests that self-generated strategies are at least as beneficial for older adults as mnemonics provided by an experimenter (Baltes, et al., 1989; Derwinger, et al., 2003; Hill, et al., 1990). Derwinger and colleagues (2005) trained older adults in a number memory task and followed up after a delay of 8 months. Differences in posttraining memory performance emerged when environmental support was withdrawn: Accuracy declined for a group that learned a mnemonic strategy, but improved for a group that had been allowed to choose their own strategies. A possible explanation for these maintenance effects is that older adults who develop their own strategies may be more likely to use them in everyday life – in effect, continuing to “train” even after the formal intervention has ended.

Furthermore, self-generated strategies are presumably more likely to be personally tailored to an individual’s strengths than a one-size-fits-all approach. Evidence from the present study supports this interpretation, showing that pre-existing verbal ability predicted training gains in the Integrated Sentences condition, but not in the Strategy Choice condition (see Figure 3.2, Row 2). This suggests that older adults with lower verbal skills were not able to derive as much benefit from an experimenter-provided encoding strategy that emphasized verbal processing. With environmental support to encourage sufficient attention and effort at encoding, however, these same older adults may have been able to make use of abilities from other domains they would not have otherwise brought to bear. In post-training questioning, successful participants from the Strategy Choice condition typically reported using deep encoding strategies, although not always involving verbal processing. Therefore, enforced study time alone appears to have been effective in directing older adults to self-initiate controlled cognitive processes.

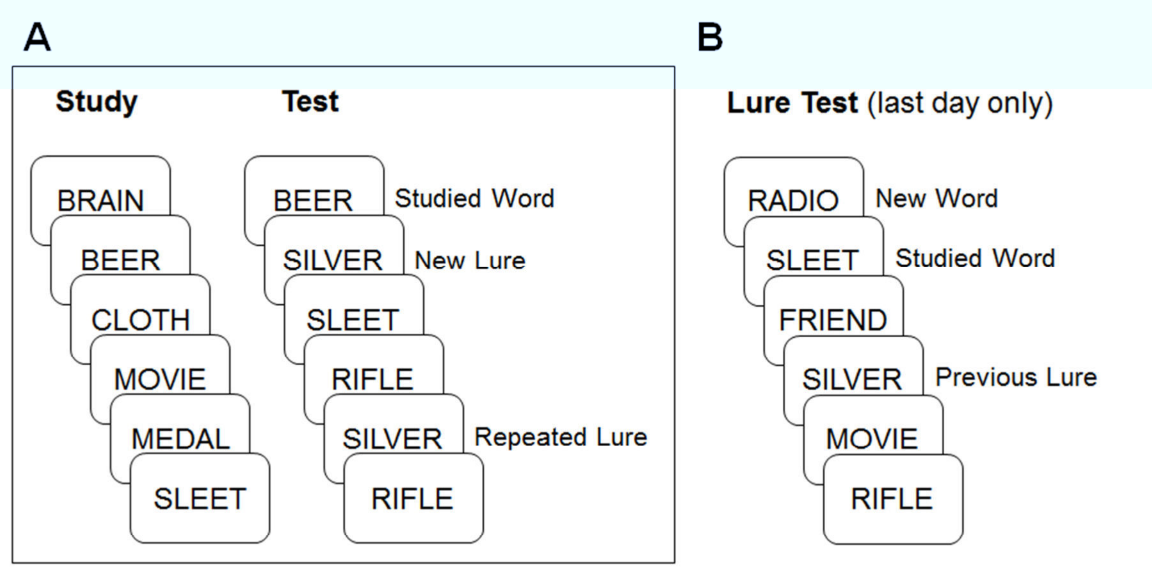
However, the influence of pre-existing ability cuts both ways. Individuals with lower initial ability levels may be most in need of intervention, yet typically show the least improvement in conventional training programs (Verhaeghen & Marcoen, 1996; Verhaeghen, et al., 1992; Yesavage, et al., 1990). In the present study, even when environmental support in the form of long encoding times was available, there was variability in the efficacy of participant-chosen encoding strategies. In post-training questioning, a subgroup of participants in the Strategy Choice condition reported using suboptimal encoding methods involving relatively shallow levels of processing, rote rehearsal, or non-specific effort-based strategies – and these participants did not benefit from training to the same extent as their peers who engaged more effective encoding methods under the very same unsupervised learning conditions. An important question for future memory training research will be how to design interventions that promote the discovery and use of successful memory strategies but allow enough flexibility for training plans to be customized to a wide range of abilities and preferences.

Acknowledgments

This chapter represents a collaboration with Cindy Lustig. We thank the many dedicated research assistants of the Lustig lab for their help on this project. We are especially grateful to Jennifer Humfleet and Caitlin Mallory for stimulus preparation and data entry, Taryn Weil, Lauren Schiff, Monica Kumar, Monika Potrykus, and Phuoc Anh Nguyen for participant scheduling, Yosi Nacach and Elyssa Guslits for coding self-report strategy data, and Katherine Cost and Mary Winters for superb lab management. We also thank the Healthier Black Elders Center at Wayne State University for their assistance with participant recruitment. Portions of this work were presented at the Annual Meeting of the Cognitive Neuroscience Society, April 2010, in Montreal.

This research was supported by the National Institute on Aging (R03AG029329) and the University of Michigan Older Americans Independence Center (P30AG024824).

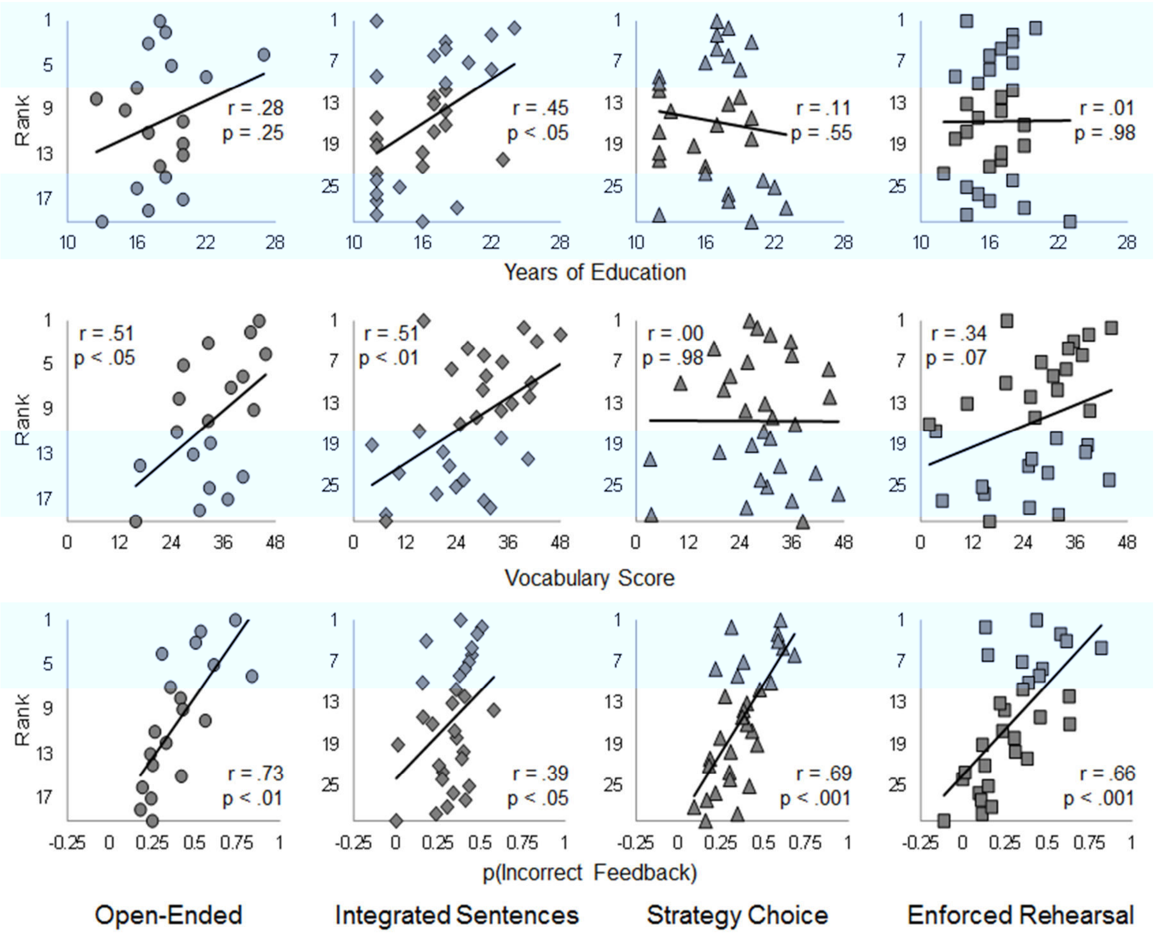
Figure 3.1 **A**: Training task. In the *study* phase, participants viewed 30 words presented individually for 14 seconds each; in the *test* phase, participants indicated whether each word was on the studied list or was an unstudied lure. Unstudied lures were repeated within each test list, with half of the repetitions occurring at a short lag (e.g., one intervening item between the first and second presentations of RIFLE in this example), and the other half occurring at a long lag (e.g., two intervening items between the first and second presentations of SILVER in this example). **B**: Surprise lure test. A surprise recognition test for unstudied lures from the final training test list was given on the last day of training; participants indicated whether each word was present in the final test session (regardless of whether it was originally a studied word or a previous lure item) or was a completely new word.



Instructions: Identify words that were on the studied list (i.e., say "**no**" to lures)

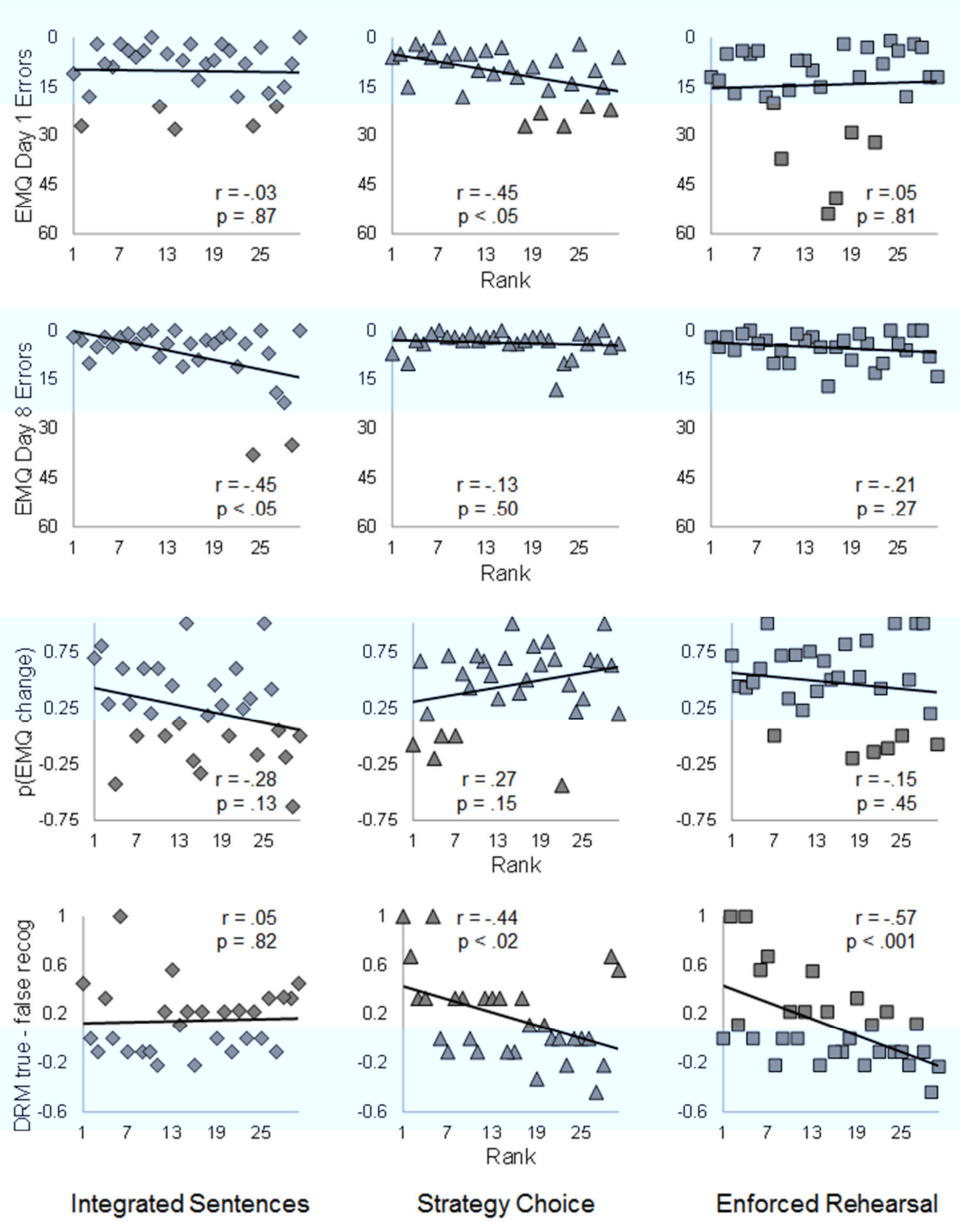
Instructions: Identify all words from the test list (i.e., say "**yes**" to lures as well as to studied words)

Figure 3.2 Correlations between training rank and ability measures, and attention to feedback after incorrect responses.



Note: *p*(Incorrect Feedback) = (incorrect feedback RT - correct feedback RT)/(incorrect feedback RT + correct feedback RT). RT = reaction time. Open-Ended = data from Bissig & Lustig (2007), in which encoding time and strategy were unconstrained.

Figure 3.3 Correlations between training rank and Everyday Memory Questionnaire scores, and DRM false memory.



Note: Axes are arranged so that better scores (e.g., fewer memory errors, greater resistance to false memories) are always higher on the y axis. *p*(EMQ change) = (Day 1 errors - Day 8 errors)/(Day 1 errors + Day 8 errors).

Table 3.1 Participant demographics by condition.

	Integrated Sentences		**Strategy Choice**		**Enforced Rehearsal**	
	M	*SD*	M	*SD*	M	*SD*
Age (years)	75.7	*6.7*	75.6	*6.7*	75.2	*7.3*
Edu. (years)	16.3	*3.7*	16.7	*3.3*	16.4	*2.4*
ERVT	27.4	*11.3*	28.6	*10.7*	26.7	*11.7*
MMSE	28.4	*1.6*	28.7	*1.5*	28.0	*1.6*
SBT	1.0	*1.5*	1.9	*2.3*	1.6	*1.6*

Note: ERVT = Extended Range Vocabulary Test (maximum score = 48); MMSE = Mini Mental State Evaluation (maximum score = 30; higher scores = better performance); SBT = Short Blessed Test (maximum score = 28; higher scores = worse performance).

<u>Table 3.2</u> Transfer task performance by condition.

	Integrated Sentences				**Strategy Choice**				**Enforced Rehearsal**			
	<u>Pretraining</u>		<u>Posttraining</u>		<u>Pretraining</u>		<u>Posttraining</u>		<u>Pretraining</u>		<u>Posttraining</u>	
	<u>M</u>	<u>*SD*</u>	<u>M</u>	<u>*SD*</u>	<u>M</u>	<u>*SD*</u>	<u>M</u>	<u>*SD*</u>	<u>M</u>	<u>*SD*</u>	<u>M</u>	<u>*SD*</u>
Face-name association recall accuracy (%)	14.7	*19.3*	22.3	*20.3*	28.3	*24.8*	37.0	*29.8*	19.0	*16.3*	28.3	*25.1*
Face-name association total accuracy (recall plus recognition; %)	89.3	*12.0*	87.7	*15.0*	88.3	*13.9*	87.3	*12.8*	91.7	*11.5*	87.7	*11.7*
Pattern Comparison Test (no. of correct responses in 20 s)	10.1	*2.3*	9.5	*2.7*	9.7	*2.1*	10.4	*2.9*	10.0	*1.5*	9.9	*2.5*
Trail-Making Test, Version A (in seconds)	37.2	*17.5*	33.9	*10.7*	40.5	*19.7*	35.8	*15.7*	33.7	*10.6*	31.9	*9.8*
Trail-Making Test, Version B (in seconds)	109.9	*86.5*	83.3	*45.6*	97.2	*46.5*	83.8	*34.7*	99.7	*55.4*	90.6	*43.1*
SOPT pattern (no. of unique responses)	11.6	*1.7*	11.5	*1.6*	11.9	*1.4*	12.0	*1.4*	11.7	*1.4*	12.2	*1.7*
SOPT word (no. of unique responses)	14.3	*1.4*	14.3	*1.5*	14.1	*1.4*	14.5	*1.4*	13.9	*1.5*	14.1	*1.5*
Shopping-list recognition accuracy (%)	95.1	*5.7*	92.8	*7.2*	96.8	*3.9*	94.8	*8.4*	96.4	*6.9*	93.4	*9.1*
Studied item accuracy (%)	93.0	*12.4*	93.6	*6.3*	95.3	*6.8*	96.0	*5.7*	94.6	*6.7*	92.1	*11.0*
Never-studied lure accuracy (%)	96.3	*6.9*	93.9	*9.5*	97.4	*4.8*	96.6	*6.6*	97.1	*8.2*	95.3	*10.5*
Previously-studied lure (from Pretraining) accuracy (%)			90.6	*10.8*			91.5	*16.9*			92.4	*11.0*
MSEQ overall memory self-efficacy strength	64.7	*18.5*	63.2	*15.9*	65.5	*19.4*	66.0	*17.7*	59.9	*19.3*	65.4	*20.3*
EMQ number of memory errors	10.2	*8.4*	7.2	*9.6*	10.7	*7.6*	3.7	*3.8*	14.4	*13.5*	5.1	*4.5*

<u>Note</u>: SOPT = Self-ordered pointing test (maximum score = 16); MSEQ = Memory Self-Efficacy Questionnaire (maximum score = 100); EMQ = Everyday Memory Questionnaire.

Table 3.3 Posttest-only task performance by condition.

	Integrated Sentences		**Strategy Choice**		**Enforced Rehearsal**	
	M	*SD*	M	*SD*	M	*SD*
Lure Test accuracy						
New items (%)	88.5	*11.8*	89.9	*6.3*	89.1	*10.1*
Studied items (%)	89.0	*12.7*	91.5	*13.2*	92.4	*12.3*
Previous lure items (%)	50.9	*34.6*	65.3	*22.4*	63.1	*22.7*
Proportion of studied items to previous-lure items	0.37	*0.08*	0.19	*0.03*	0.21	*0.04*
DRM False Memory Test accuracy (%)						
Critical lures	20.2	*24.6*	25.5	*29.9*	18.8	*29.9*
Studied list items	90.6	*11.9*	93.0	*10.2*	93.0	*8.9*
Unstudied list items	91.0	*12.0*	96.3	*7.8*	91.9	*13.3*
Unstudied list lures	87.0	*24.5*	97.8	*8.4*	93.4	*16.1*
DRM False Memory Test: corrected true recognition - corrected false recognition [a]	14.8	*27.4*	17.0	*35.6*	10.4	*35.2*
Word memory accuracy (%)	71.8	*9.9*	76.5	*10.4*	79.3	*9.7*
Source memory accuracy (%)	50.6	*17.1*	52.9	*15.8*	53.7	*16.5*

[a]: Lower scores indicate greater susceptibility to false memories; higher scores indicate greater resistance to false memories.

CHAPTER 4

STABILITY AND LABILITY IN FALSE RECOGNITION: EFFECTS OF ENCODING MANIPULATIONS AT SHORT AND LONG DELAYS

Introduction

The susceptibility of memory to distortion has been empirically studied using the converging associates or DRM paradigm (Deese, 1959; Roediger & McDermott, 1995), in which unstudied semantic associates (“related lures”) to lists of studied words are later endorsed as studied items. The ease with which false memories can be reliably induced using the DRM procedure is impressive, but it also invites the critical question of how such memory errors can be reduced. By extension, studying factors that increase and decrease the incidence of false memories in the laboratory has the potential to positively affect memory accuracy in daily life. Previous DRM studies have tested the effects of varying word list construction (McDermott, 1996; Robinson & Roediger, 1997), encoding instructions (Rhodes & Anastasi, 2000; Thapar & McDermott, 2001; Toglia, et al., 1999), testing conditions (Coane & McBride, 2006; McDermott, 1996), and even providing explicit warnings about the nature of the word lists and related lures (Anastasi, et al., 2000; Gallo, Roberts, & Seamon, 1997; McDermott & Roediger, 1998; Neuschatz, et al., 2001). The persistent finding is that while false memory rates can be attenuated through such manipulations, they are rarely if ever eliminated.

From both theoretical and applied perspectives, a relevant question is whether false memories are more influenced by processes involved at encoding or retrieval. Identifying the origin of memory distortions would greatly help in targeting interventions to minimize their occurrence. On this issue, some researchers favor memory-based explanations, which claim that memory traces can be enriched at encoding (e.g., by increasing attention to distinctive characteristics) in ways that will later support accurate recall or recognition (Arndt & Reder, 2003; McCabe, et al., 2004; Smith & Hunt, 1998). Other researchers favor decision-based explanations, which state that criteria can be adjusted at retrieval (e.g., by increasing expectation of distinctive characteristics) to increase memory accuracy (Dodson & Schacter, 2002b; Schacter, Cendan, Dodson, & Clifford, 2001). In the DRM literature, as reviewed below, evidence thus far has been largely equivocal (Arndt, 2006; Hege & Dodson, 2004).

According to the levels-of-processing theory (Craik & Lockhart, 1972), information encoded at a *deep*, semantic level will be better remembered than information encoded at a *shallow*, perceptual level. However, research using the DRM paradigm has shown that deep processing of semantically associates (e.g., making pleasantness ratings, or concrete/abstract decisions) increases the rate of false recall (Rhodes & Anastasi, 2000; Toglia, et al., 1999) and false recognition (Thapar & McDermott, 2001) of related lure words, as well as increasing accurate memory for studied items. “Fuzzy trace” theory (Brainerd &

Reyna, 2002; Reyna & Brainerd, 1995) offers an explanation for these results, based on the claim that two types of information are encoded into memory in parallel: verbatim, item-specific traces which faithfully record surface features and details of an experience, and gist representations which store general themes or semantic content. For accurate memory of semantically-related material, verbatim and gist traces reinforce each other, but gist traces also promote false memory for related information, especially when unopposed by short-lasting verbatim traces. Deep, meaning-based encoding strengthens durable gist traces, accounting for the resultant increases in both accurate memory and false memory observed in DRM studies.

Within this theoretical framework, a potential tactic for reducing false memory rates would be to encourage processing that selectively strengthens verbatim traces, and reduces the availability of gist traces. This approach has been tested in the DRM literature by contrasting *item-specific* processing, which emphasizes unique characteristics that differentiate studied items, with *relational* processing, which emphasizes similar characteristics that studied items have in common (Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993). Both item-specific and relational encoding support accurate memory for semantically-related material, but only item-specific information aids in rejecting false memories. Some researchers have varied the distinctiveness of stimulus features in order to induce item-specific processing, e.g., by presenting memoranda as pictures rather than words (Hege & Dodson, 2004; Israel & Schacter, 1997; Schacter, et al., 1999) or by presenting words in unusual fonts

(Arndt, 2006; Arndt & Reder, 2003); other researchers have manipulated encoding instructions (Butler, et al., 2010; McCabe, et al., 2004) to encourage or discourage item-specific processing. Importantly, past studies have also varied in the use of within-subjects versus between-subjects designs. A consistent finding is that item-specific processing is associated with reductions in false memories, but there is debate over whether the protective effect originates at encoding (Arndt & Reder, 2003; McCabe, et al., 2004; Smith & Hunt, 1998) or retrieval (Dodson & Schacter, 2002b; McCabe & Smith, 2006; Schacter, et al., 2001).

Decision-based explanations (the most prominent of which is the "distinctiveness heuristic"; Schacter, et al., 1999) posit that false memories are averted by applying a global strategy at retrieval, in which distinctive information is expected for studied items that were perceived as distinctive at encoding. Because such a strategy is only effective in between-subjects designs (participants cannot use the presence or absence of distinctive information as a valid decision criterion if they experienced both high-distinctiveness and low-distinctiveness stimuli at encoding; Dodson & Schacter, 2001; Schacter, et al., 1999), finding decreased false memory rates in a within-subjects design would instead favor memory-based explanations, which hold that item-specific processing strengthens verbatim memory traces at encoding while reducing the generation (and later availability) of gist traces (Arndt & Reder, 2003; McCabe, et al., 2004; Smith & Hunt, 1998). In fact, several recent investigations have found support for contributions of item-specific processing at both encoding and retrieval (Arndt, 2006; Hege & Dodson, 2004). Because the integrity of verbatim traces rapidly declines, manipulating encoding instructions and probing for false memories at short and long delays may reveal more about the mechanisms of item-specific processing.

We recently developed a novel experimental paradigm to directly compare false memories in the same participants under short-term memory (STM) and long-term memory (LTM) conditions (Flegal, Atkins, & Reuter-Lorenz, 2010). This research has followed on from initial demonstrations of false recognition and false recall with short lists of semantically-related words after brief delays, which extended LTM false memory effects into the domain of STM, and exposed the surprising limitations of verbatim retention (Atkins & Reuter-Lorenz, 2008). A unique advantage to the design of our new task is its ability to dissociate retrieval processes operating under STM and LTM conditions, for items studied in the same encoding period – but which may be crossed with multiple encoding strategies. In our previous study (Flegal, et al., 2010), no encoding strategy instructions were given. Participants viewed lists of four semantic associates and were probed immediately following a filled 3-4 second retention interval (i.e., STM) or approximately 20 minutes later in a surprise recognition test (i.e., LTM). Measurements of both quantity (the relative incidence of false recognition) and quality (the accompanying phenomenology, or subjective experience) of memory distortions did not reliably differ between short-term and long-term testing, contrary to predictions based on conventional models of multiple memory systems which separate STM and LTM processes.

The results from Flegal, Atkins, and Reuter-Lorenz (2010) support unitary models of memory in which STM and LTM rely on common, delay-invariant processes, and are compatible with theoretical accounts implicating variable strengths of verbatim and gist memory representations (Brainerd & Reyna, 2002; Reyna & Brainerd, 1995). Hypotheses from “fuzzy trace” theory can be further tested by assessing the effects of encoding manipulations on false memories at short as well as long delays. In a levels-of-processing framework, deep, meaning-based encoding is predicted to strengthen verbatim traces which will remain briefly available to oppose false STM, but also enhance durable gist traces which will increase false LTM. Because gist traces support both accurate memory and false memory for semantically related lists, deep encoding should likewise increase accurate memory, especially at long delays when verbatim representations (which can be strengthened by surface-based encoding) will be highly degraded. Item-specific and relational encoding can both be considered forms of “deep” processing, and thus accurate memory should benefit from either strategy, but their effects on false memory are predicted to differ. Item-specific processing should suppress false memory (but increase accurate memory) at both STM and LTM, while relational encoding should increase false memory (as well as accurate memory), particularly at LTM. According to memory-based explanations, these effects will occur because item-specific encoding reduces the availability of gist traces, even at LTM, while relational encoding strengthens durable gist traces to the exclusion of item-specific information that might oppose false memory. While there are theoretical grounds for each of these predictions, there is little experimental precedent on which to base hypotheses about encoding manipulations specifically in STM.

Rose, Myerson, Roediger, and Hale (2010) conducted one study that examined the effects of encoding manipulations at both immediate (i.e., STM) and delayed (i.e., LTM) testing, using in a novel “levels-of-processing span” paradigm. To-be-remembered words were presented serially, each one followed by a pair of “processing” words which prompted participants to make a judgment based on visual, phonological, or semantic similarity to the preceding word. At the end of each word list, an immediate recall test required participants to report the to-be-remembered words in order. After three blocks of STM trials (one for each levels-of-processing condition), LTM accuracy was assessed with a surprise delayed recognition test. Results from the two tests dissociated the effects of levels of processing on STM and LTM performance. Immediate recall accuracy was not influenced by the processing manipulation, but a classic levels-of-processing effect emerged on the delayed recognition test, where accuracy was highest for words that received deep (semantic) encoding, and lowest for words that received shallow (visual) encoding. These data challenge strict models of unitary memory systems, by demonstrating the involvement of different processes under STM and LTM conditions. However, while the Rose et al. (2010) study was informative about levels-of-processing effects on *accurate* memory at short and long delays, it is unclear whether the incidence and phenomenology of *false* memory would be similarly influenced.

The present report addresses the question of whether encoding manipulations that affect false memory will operate similarly in STM and LTM, or if their effects will dissociate these systems. Experiment 1 uses a levels-of-processing manipulation (Craik & Lockhart, 1972), and Experiment 2 compares item-specific and relational encoding (Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993). Combined with our previous study (Flegal, et al., 2010), this approach aims not only to help adjudicate between competing theoretical explanations of false memories, but also to further clarify similarities and differences between short-term and long-term memory processes.

Experiment 1

Method

Participants

32 individuals (18-24 yrs old; *M* = 19.8) participated for course credit or payment. Eight additional participants were tested but excluded from analysis for noncompliance or computer malfunction. Research protocols were approved by the University of Michigan Institutional Review Board, and all participants provided written informed consent.

Materials

The memory sets were 96 lists of four semantically related words, all associates of a common theme word. The theme word (e.g., SLEEP for the list containing the associates nap, doze, bed, awake) served as the probe on every trial. The three probe types were: *related lure*, the unstudied theme word associated with a studied list; *unrelated lure*, an unstudied theme word associated with a nonpresented list; and *target*, the theme word associated with, and present in, a studied list (replacing one of the four associates; ordinal position of the replaced item balanced across target probe trials). Theme words and probes appeared only once during the experiment; no list was probed in both STM and LTM trials.

The 96 lists were divided into four groups of 24 four-word lists (Groups A-D) equated in mean backward associative strength (*M* = 0.40). Probe type was counterbalanced with word lists across participants, so that for one quarter of all participants, lists in Group A were paired with related lure probes, B with unrelated lure probes, and C with target probes. Theme words associated with Group D lists served as the unrelated lure probes.

Each of the four groups was further divided into two subgroups of 12 four-word lists, following the same parameters, to balance the status of each list as short-term versus long-term memoranda across participants. Thus, for half of the participants in each of the four orders, the first subgroup of lists from each group (e.g., A1) was probed during the STM trials, and the second subgroup of lists (e.g., A2) during the LTM trials; the assignment was reversed for the other half of the participants. This procedure ensured that all participants encountered the same probes – all theme words, but in different contexts – as related lures, unrelated lure probes, or target probes, and as STM or LTM probes (cf. Flegal, et al., 2010).

Finally, the blocked order of encoding instructions was counterbalanced across participants, so that half of the participants in each of the orders experienced an ABBA design beginning with “shallow” encoding instructions, and the other half experienced a BAAB design beginning with “deep” encoding instructions. All participants completed a total of four STM blocks, two with each set of encoding instructions.

Design and procedure

As previously described (Flegal, et al., 2010), the four-word memory sets were probed either within the same trial (i.e., STM) or in a surprise recognition test following completion of all STM trials (i.e., LTM), in order to examine short-term and long-term memory distortions concurrently and within-subjects (see Figure 4.1). As in our previous study, each STM trial started with a brief study period followed by a 3,000-4,000 ms retention interval filled with a math equation verification task, then (a) the probe word to respond to (yes or no) and a four-point confidence rating (*very low*, *somewhat low*, *somewhat high*, or *very high*), or (b) cues for two arbitrary button presses with corresponding response mappings (for memory sets that would subsequently be probed at LTM). Instead of simultaneous presentation of the entire memory set at the beginning of each STM trial, however, here the four words were presented serially (for 250 ms each, with a 2,000 ms inter-stimulus interval) to allow for individual encoding judgments.

Shallow encoding instructions directed participants to “count the total number of ascenders and descenders in each word,” thereby requiring attention to the morphological characteristics of the memory set items (always presented in lower-case letters). Participants made a four-point response to each of the four words on these trials, indicating the sum of its ascenders and descenders: 0, 1, 2, or 3+. Ascenders (b, d, f, h, k, l, and t) and descenders (g, j, p, q, y) were defined for participants in a practice session prior to the start of the experiment.

<u>Deep encoding instructions</u> directed participants to “Rate how much you like the meaning of each word,” thereby requiring attention to the semantic characteristics of the memory set items. Participants made a four-point liking response to each of the four words on these trials: dislike a lot, dislike a little, like a little, or like a lot.

The STM trials consisted of four blocks, two for each set of encoding instructions, presented in ABBA order. Instructions for both encoding conditions were given before starting the first STM block. Half of the participants began with a shallow encoding block and the other half began with deep encoding block. All participants received a 1-min break between the second and third blocks. Each block of trials started with a reminder screen describing the encoding instructions and response options, and an identifying colored border (shallow = blue; deep = pink) remained on the screen throughout the block. Participants completed 18 trials in each of the 4 blocks, presented in random order. Of the 9 trials in each block probed at STM, 3 were of each probe type (related lure, unrelated lure, and target).

A 2-min break followed completion of the STM trials, then participants were informed about, and given instructions for, the LTM recognition test. Each participant completed 72 LTM trials, 36 of which tested memory sets that were not probed at STM (12 each – 3 from each block – of the related lure, unrelated lure, and target probe type). Additionally, there were 12 trials of studied associates from memory sets that were probed at STM (never including theme words from target probe trials), and 24 trials of unstudied, unrelated foils, matched for frequency and word length with the corpus of theme words used in the experiment. Each LTM trial started with the probe word to respond to (by indicating whether or not it had appeared during the STM trials: *yes* or *no*) followed by a four-point confidence rating.

Results and Discussion

Encoding time

The mean response latency to memory set items in the STM trials (i.e., the amount of time to make encoding judgments) was significantly longer in the shallow, “letters” encoding condition (*M* = 1,001 ms, *SEM* = 26 ms) than in the deep, “liking” encoding condition (*M* = 809 ms, *SEM* = 26 ms), $t(31) = 5.69$, $p < 0.001$. Although this may indicate systematic differences in the average amount of encoding time, it should be noted that the memory set items were present on the computer screen for the same duration (250 ms each) in each condition; they did not remain visible only until a response was made. The difference is also in the opposite direction from that which might predict better memory associated with deeper encoding and longer encoding time: response times were actually faster for liking ratings (deep) than letter counting (shallow).

Accuracy

Mean math task accuracy during the STM trial retention interval was 0.89; it did not significantly differ between shallow encoding (*M* = 0.89) and deep encoding (*M* = 0.88) blocks. As shown in Figure 4.2, a classic levels-of-processing effect was evident for studied items, as correct recognition of target probes in the deep encoding condition remained nearly as high on LTM trials (*M* = 0.84) as on STM trials (*M* = 0.87), compared a steep drop in target probe accuracy from STM (*M* = 0.82) to LTM (*M* = 0.53) in the shallow encoding condition. The encoding condition (shallow, deep) × delay (STM, LTM) interaction for target probe accuracy was highly significant, *F*(1,31) = 21.40, *p* < .001.

At short-term testing, there was a significant false memory effect (the difference between *yes* responses to unstudied, related lure probes and unstudied, unrelated lure probes) for both shallow encoding (*M* = 0.16) and deep encoding (*M* = 0.11). At long-term testing, however, the false memory effect no longer reliably differed from zero for lists that received shallow encoding (*M* = 0.04), while the false memory effect for lists that received deep encoding was more than double in size at LTM (*M* = 0.27) compared to STM (see Figure 4.2). Thus, although the incidence of false recognition did not differ between encoding conditions at STM (*t*(31) = 0.98, *p* = .33), it was significantly greater in the deep encoding than the shallow encoding condition at LTM (*t*(31) = 2.96, *p* < .01). The finding that deep, sematic processing at encoding increased rates of both accurate and false memory at long-term testing is consistent with prior results from the DRM literature (e.g., Thapar & McDermott, 2001).

The baseline (unrelated lure probe) false alarm rate at LTM was equivalent for shallow (*M* = 0.28) and deep (*M* = 0.26) encoding conditions, indicating that the significant difference in the false memory effect at LTM was driven by a disproportionate increase in false alarms to related lure probes belonging to lists that received deep encoding at STM. This time-dependent effect of levels of processing on false memory formation was evident in a significant lure type (related, unrelated) × encoding condition (shallow, deep) × delay (STM, LTM) interaction, *F*(1,31) = 10.16, *p* < .01. While our earlier study (Flegal, et al., 2010) found relatively delay-invariant false memory effects when encoding strategies were unconstrained, the present results imply that there may still be different mechanisms underlying memory formation and monitoring in STM and LTM.

In summary, the false memory effect in Experiment 1 decreased over time for lists that were studied under shallow encoding instructions (from *M* = 0.16 to *M* = 0.04), but increased over time for lists that were studied under deep encoding instructions (from *M* = 0.11 to *M* = 0.27). Notably, the finding of a stable false memory effect from short-term to long-term testing, as reported in Flegal, Atkins, and Reuter-Lorenz (2010), is replicated when the present data are collapsed across encoding condition (as shown by a non-significant lure type (related, unrelated) × delay (STM, LTM) interaction: *F*(1,31) = 0.23, *p* = .64). This suggests that in our previous study, where encoding instructions were not experimentally controlled, there was likely considerable variability in processing strategies engaged across participants, ranging from “shallow” to “deep”.

Phenomenological experience

The finding of stable levels of confidence in false recognition over time, as reported in Flegal, Atkins, and Reuter-Lorenz (2010), was replicated in both encoding conditions. As shown in Figure 4.3, confidence in “yes” responses to related lures did not change from STM to LTM for lists that received shallow encoding (*t*(13) = 1.38, *p* = .19) or lists that received deep encoding (*t*(14) = 0.49, *p* = .63). In contrast, but also consistent with results from our earlier study under unconstrained encoding instructions, confidence in “no” (correct) responses to related lures significantly decreased from short-term to long-term testing in both the shallow (*t*(30) = 3.43, *p* < .01) and deep (*t*(30) = 3.37, *p* < .01) encoding conditions. The response (*yes*, *no*) × delay (STM, LTM) interaction for related lure confidence ratings was marginally significant in both the shallow (*F*(1,12) = 3.92, *p* = .07) and deep (*F*(1,14) = 4.26, *p* = .06) encoding conditions. Thus, the effects of levels of processing did not dissociate STM and LTM on participants’ confidence in their responses to related lure probes. For all correct responses, however, there was a significant main effect of encoding condition (*F*(1,29) = 33.87, *p* < .001), such that confidence ratings were higher overall for lists that were studied under deep than under shallow encoding instructions.

To summarize, accuracy data from Experiment 1 revealed time-dependent effects of processing depth on false memory formation. A direct comparison of short-term and long-term testing indicates that deep encoding preserved accurate memory across delay and increased false memory at LTM (consistent with results from DRM studies; e.g., Thapar & McDermott, 2001), while the encoding strategy manipulation had minimal effects at STM (consistent with results from the levels-of-processing span task; Rose, et al., 2010). Confidence data replicated our earlier findings of statistically equivalent measurements of false memory phenomenology at short and long delays (Flegal, et al., 2010), which did not differ as a function of encoding strategy in the present study.

Because evidence suggests that deep processing strengthens both verbatim and gist memory traces, independent contributions of the two types of memorial information are difficult to discern in Experiment 1. Especially under STM conditions, when verbatim traces are still available, it may be that the false memory effect was not greater for lists studied under deep than shallow encoding instructions because an enhanced gist representation was canceled out by an enhanced verbatim representation. To better understand the processes responsible for memory distortions, Experiment 2 tests for time-dependent effects with an encoding manipulation designed to selectively strengthen verbatim and gist memory traces (through item-specific and relational processing, respectively). Experiment 2 also implements Remember/Know judgments (Gardiner, 1988; Tulving, 1985; Yonelinas, 2002), in an effort to better capture differences in false memory phenomenology than the confidence ratings used in Experiment 1. Furthermore, a within-subjects design (Experiment 2A) is compared to a between-subjects design (Experiment 2B) to address the possibility of order effects on encoding instruction adherence, and to test competing predictions from memory-based and decision-based explanations about the effects of item-specific processing on false memory.

Experiment 2A

Method

Participants

32 individuals (18-21 yrs old; M = 18.6) participated for course credit. Six additional participants were tested but excluded from analysis for a) recognition accuracy scores > 2.5 standard deviations from the mean, or b) post-experiment questionnaire responses indicating they failed to understand the Remember/Know distinction.

Design and procedure

The method was the same as Experiment 1 except that a different pair of encoding instructions was used (described below), and remember/know/guess judgments replaced confidence ratings. Following each "yes" response to a probe word, participants used their right hand to indicate whether they *remember* the probe word was in the memory set (recollecting something distinctive about studying the word), they *know* the probe word was present (recognizing the word without retrieving specific details of its study), or their response had been a *guess*. As in our previous study (Flegal, et al., 2010), detailed instructions explaining the Remember/Know distinction were adapted from Rajaram (1993). To equate the number of responses on each trial, a display of three boxes appeared following each "no" response to a probe word, prompting an arbitrary right-handed response.

Furthermore, in Experiment 2, the blocked order of encoding instructions was counterbalanced across participants so that instructions for the first half of

STM trials were given before the first block, and instructions for the second half of STM trials were given between the second and third blocks. Half of the participants in each of the orders experienced an AABB design beginning with “item-specific” encoding instructions, and the other half experienced a BBAA design beginning with “relational” encoding instructions. All participants completed a total of four STM blocks, two with each set of encoding instructions. Unlike Experiment 1, no manual responses (corresponding to individual encoding judgments) were required for memory set items in the STM trials.

Item-specific encoding instructions directed participants to “think of a characteristic that makes each word *unique* in a way that differentiates it from other words.” Relational encoding instructions directed participants to “think of a characteristic that makes each word *similar* to other, related words.” Detailed instructions, with examples of unique and similar characteristics, were adapted from Butler, McDaniel, McCabe, & Dornburg (2010).

Results and Discussion

Accuracy

Mean math task accuracy during the STM trial retention interval was 0.89; it did not significantly differ between item-specific encoding (*M* = 0.89) and relational encoding (*M* = 0.89) blocks. Significant false memory effects were present at both delays, for both encoding conditions, as participants made more false alarms to related lure probes than to unrelated lure probes (Table 4.1A). The size of the false memory effect at STM was greater for lists studied under item-specific encoding instructions (*M* = 0.18) than under relational encoding instructions (*M* = 0.10), which was a marginally significant difference (*t*(31) = 1.84, *p* = .08), but the false memory effect was equivalent at LTM (*M* = 0.27 for both conditions) and the lure type (related, unrelated) × encoding condition (shallow, deep) × delay (STM, LTM) interaction was not significant, *F*(1,31) = 1.18, *p* = .29. Collapsed across encoding condition, a significant lure type × delay interaction indicated that the size of the false memory effect was greater overall at LTM than STM (*F*(1,31) = 14.97, *p* < .001), unlike Experiment 1.

The finding of larger false memory rates at STM for the item-specific encoding condition that the relational encoding condition was unexpected. Furthermore, predicted reductions in false memory for item-specific compared to relational encoding at long-term testing, based on prior results from the DRM literature (e.g., McCabe, et al., 2004), were not observed. It is possible that order effects were contaminating the present data, in that participants who received one set of encoding instructions first (i.e., for the first two STM blocks) may have been more likely to use that strategy as instructed than participants who received that set of encoding instructions second (i.e., after exposure to the other set of encoding instructions). Because the item-specific and relational processing strategies required no overt response at encoding (in contrast to the letter-counting and liking-rating responses collected in Experiment 1), adherence to instructions could not be measured directly. Visual inspection of the relevant means gives some indication that for participants who received relational encoding instructions first, the false memory effect was less influenced by the change in encoding strategies than for participants who received item-specific encoding instructions first (Table 4.1A). However, an ANOVA including the between-subjects factor of order (item-specific first, relational first) yielded no significant main effects or interactions.

Phenomenological experience

An ANOVA on normalized estimates of false recollection, the proportion of *remember* responses to related lures out of the total proportion of *yes* responses to related lures, showed no significant main effect of delay ($F < 1$). Thus, the subjective experience of “remembering” an item that had never been studied appeared equally robust at STM and LTM, collapsed across encoding condition. However, a significant main effect of encoding condition ($F(1,6) = 7.19$, $p < .05$), revealed a higher incidence of *remember* responses associated with false recognition for lists studied under relational encoding instructions than under item-specific encoding conditions, collapsed across delay.

These results replicate findings of stability in the normalized incidence of *remember* responses to related lures from STM to LTM, as reported in Flegal, Atkins, and Reuter-Lorenz (2010), and are consistent with similar effects found in confidence ratings from Experiment 1 in the present study, but they should be interpreted with caution because relatively few participants assigned *remember* responses to related lure false alarms, especially at short-term testing, and thus the sample size was restricted to $n = 7$ for the delay (STM, LTM) × encoding condition (item-specific, relational) overall ANOVA. Paired *t* tests at each delay did not show statistically significant differences between the encoding conditions at STM (*t*(6) = 1.80, *p* =.12) or at LTM (*t*(23) = 1.38, *p* = .18), although the normalized incidence of *remember* responses to related lures was numerically greater in the relational encoding condition than the item-specific encoding condition at both short and long delays.

Experiment 2B

Method

Participants

28 individuals (18-21 yrs old; *M* = 18.9) participated for course credit. Six additional participants were tested but excluded from analysis, following the same criteria as Experiment 2A.

Design and procedure

The method was identical to Experiment 2A except the encoding strategy manipulation was between-subjects instead of within-subjects. Half of the participants completed all of the STM trials under item-specific encoding instructions, and the other half completed all of the STM trials under relational encoding instructions.

Results and Discussion

Accuracy

Mean math task accuracy during the STM trial retention interval was 0.89; it did not significantly differ between item-specific encoding (*M* = 0.88) and relational encoding (*M* = 0.90) conditions. Significant false memory effects were present at both delays, for both encoding conditions (Table 4.1B). A lure type (related, unrelated) × delay (STM, LTM) × encoding condition (between subjects: item-specific, relational) mixed design ANOVA revealed that the size of the false memory effect was greater overall at LTM than STM (*F*(1,26) = 4.56, *p* < .05), but there was no significant main effect of encoding condition (*F*(1,26) = 1.78, *p* = .19) or three-way interaction (*F* < 1). As in Experiment 2A – and although these data were free from the confound of prior exposure to another encoding strategy – the size of the false memory effect at STM was numerically (but not reliably; *p* = .44) greater for lists studied under item-specific encoding instructions (*M* = 0.16) than under relational encoding instructions (*M* = 0.13), and the false memory effect was equivalent at LTM (*M* = 0.24 and *M* = 0.25).

<u>Phenomenological experience</u>

An ANOVA on normalized estimates of false recollection showed no significant main effect of delay (*F* < 1). Thus, as in Experiment 2A, the proportion of *remember* responses to related lures out of the total proportion of *yes* responses to related lures was stable from STM to LTM, collapsed across encoding condition. Unlike Experiment 2A, however, here there is no significant main effect of the between-subjects factor of encoding condition (*F* < 1), collapsed across delay. Similarly infrequent rates of *remember* responses to related lure false alarms mean that these results must also be interpreted with caution. At STM, the normalized incidence of *remember* responses to related lures did not differ between encoding conditions (*t*(14) = 0.19, *p* = .85), and although the difference at LTM was not reliable (*t*(25) = 0.98, *p* = .34), the proportion of trials on which related lures were falsely “remembered” was numerically greater for lists studied under relational encoding instructions (*M* = 0.30) than lists studied under item-specific encoding instructions (*M* = 0.18).

Taken together, the results from Experiments 2A and 2B are difficult to reconcile with previous DRM studies on item-specific and relational processing, for a number of reasons. Both encoding strategies can be considered forms of “deep” processing, thus predicting that accurate memory should benefit from either strategy, and indeed we did not observe differential effects of encoding condition on target probe accuracy in either the within-subjects or between-subjects design (Table 4.1). However, there is no theoretical basis on which to predict that item-specific processing would produce *higher* rates of false STM than relational processing, and yet this was a consistent finding in the present data. The size of the false memory effect at short-term testing was numerically greater for lists studied under item-specific encoding instructions than under relational encoding instructions, both when the manipulation was within-subjects (Experiment 2A) and between-subjects (Experiment 2B). Additionally, theoretical considerations and results from previous DRM studies supported the prediction that relational processing should increase rates of false LTM, relative to item-specific processing. On the contrary, the size of the false memory effect at long-term testing did not significantly differ between item-specific and relational encoding conditions, whether the manipulation was within-subjects (Experiment 2A) or between-subjects (Experiment 2B). While relational processing did not produce the expected increases in false memory either under STM or LTM conditions, it was associated with a higher normalized incidence of *remember* responses to related lures. Although the results must be considered tentative because they are based on small sample sizes, these effects suggest that strengthening gist traces through relational encoding enhances the illusory qualities of false memories, even at short delays. The relative time-invariance of these effects, together with the same pattern of stability in lure confidence ratings from Experiment 1, provides further evidence for similar phenomenology of semantic distortions at short-term and long-term testing, implicating a common processing basis for these forms of false recognition.

General Discussion

In Experiment 1, a levels-of-processing manipulation had little effect on recognition performance at short delays, but deep encoding (relative to shallow encoding) elevated the rates of both accurate and false memory at longer delays. The finding that effects from an encoding strategy manipulation at STM emerged only at LTM testing is similar to the dissociation that Rose and colleagues (2010) observed in their levels-of-processing span task. The direction of the deep processing effects that appeared at LTM in the present study (i.e., increases in false recognition as well as accurate recognition) also reproduces published findings (Thapar & McDermott, 2001), demonstrating that encoding influences on false LTM can originate even from four-word lists of semantic associates studied under STM conditions.

While the effects of levels of processing on false memory were found to be time-dependent, the underlying incidence and phenomenological qualities of false memory were relatively time-invariant, as we have previously shown. Collapsed across encoding condition, the size of the false memory effect in Experiment 1 did not differ between short-term and long-term testing. Similarly, self-reports of confidence in false alarms to related lures were statistically equivalent across delay. These data replicate the results reported in Flegal, Atkins, and Reuter-Lorenz (2010), further supporting the claim that compelling false memory illusions can arise within seconds of encoding.

Experiment 2 used a different encoding manipulation (item-specific versus relational processing) and produced several unexpected results. At STM, item-specific processing was associated with higher rates of false recognition than relational processing, contrary to the prediction that item-specific processing would suppress false memory. At LTM, the false memory effect for lists studied under item-specific encoding instructions did not differ from lists studied under relational encoding instructions, contrary to the prediction that relational processing would increase false memory. Initial results from Experiment 2A suggested that the order in which encoding instructions were received may have confounded the encoding strategy manipulation, but the same pattern of effects was then observed with a between-subjects design in Experiment 2B.

It is not clear why the effects of item-specific and relational processing on false memory formation did not differ between within-subjects and between-subjects designs in the present study. If item-specific processing had been found to decrease false memory rates in a within-subjects design, such evidence would support memory-based explanations of false memories (Arndt & Reder, 2003; McCabe, et al., 2004; Smith & Hunt, 1998), because decision-based explanations require a between-subjects manipulation of the availability of distinctive information (Dodson & Schacter, 2001; Schacter, et al., 1999). However, no benefit of item-specific encoding was found in either the within-subjects design or the between-subjects design used in Experiment 2. Future research will be necessary to determine the cause for this null effect. Because no overt response was required at encoding for item-specific and relational processing, the level of adherence to instructions may not have been high. Additionally, Experiment 2 tested the effects of a processing manipulation by varying encoding instructions, but an alternative approach to consider involves varying the perceptual distinctiveness of stimuli in order to promote item-specific processing – a manipulation that has been found effective for reducing false memory rates in the domain of LTM (e.g., Arndt & Reder, 2003; Israel & Schacter, 1997).

In contrast to the converging evidence for process constancy in Experiment 1 and our earlier study (Flegal, et al., 2010), when recognition accuracy data were collapsed across encoding condition in Experiment 2 the false memory effect increased significantly from STM to LTM. This may have been a consequence of deep, semantic processing (required by both item-specific and relational encoding instructions) creating a main effect of larger false memory effects at later delays – whereas a range of encoding strategies, from shallow to deep, was being employed across participants in the other studies which showed delay-invariant false memory effects. However, stability over time was observed in Experiment 2 in the proportion of false memories associated with “remember” phenomenology. This effect was present in both the within-subjects and between-subjects designs, it mirrors the finding of stable levels of confidence in lure false alarms from Experiment 1, and it also replicates the same finding (statistical equivalence in the proportion of *remember* responses to related lures out of the total proportion of *yes* responses to related lures at STM and LTM) that was present in our previous study under unconstrained encoding conditions. Although difficult to interpret in the context of unpredicted effects of the item-specific and relational encoding manipulations on other dependent measures, this effect illustrates the robustness of false memory illusions and adds to a body of evidence indicating that the processes that give rise to memory distortions may be the same at short and long delays.

Acknowledgments

This chapter represents a collaboration with Patricia A. Reuter-Lorenz. We thank Halle Zucker, Puneet Sodhi, Ashley Brooks, Michael Kinning, Emily Sallen, Greta Bielaczyc, Alexander Kapordelis, Phebe Brenne, Amanda Perez, Gabriela Ramirez, Elizabeth Irish, Samantha Winter, Teija Madhusoodanan, and Sarah Koltun for their assistance with data collection, and we extend special thanks to Hiten Patel for his extensive contributions to Experiment 2.

Figure 4.1 Experiment 1 design. Each four-word list is probed only once; either immediately following a 3-4 second filled retention interval (short-term memory), or in a surprise recognition test after all lists are encoded (long-term memory).

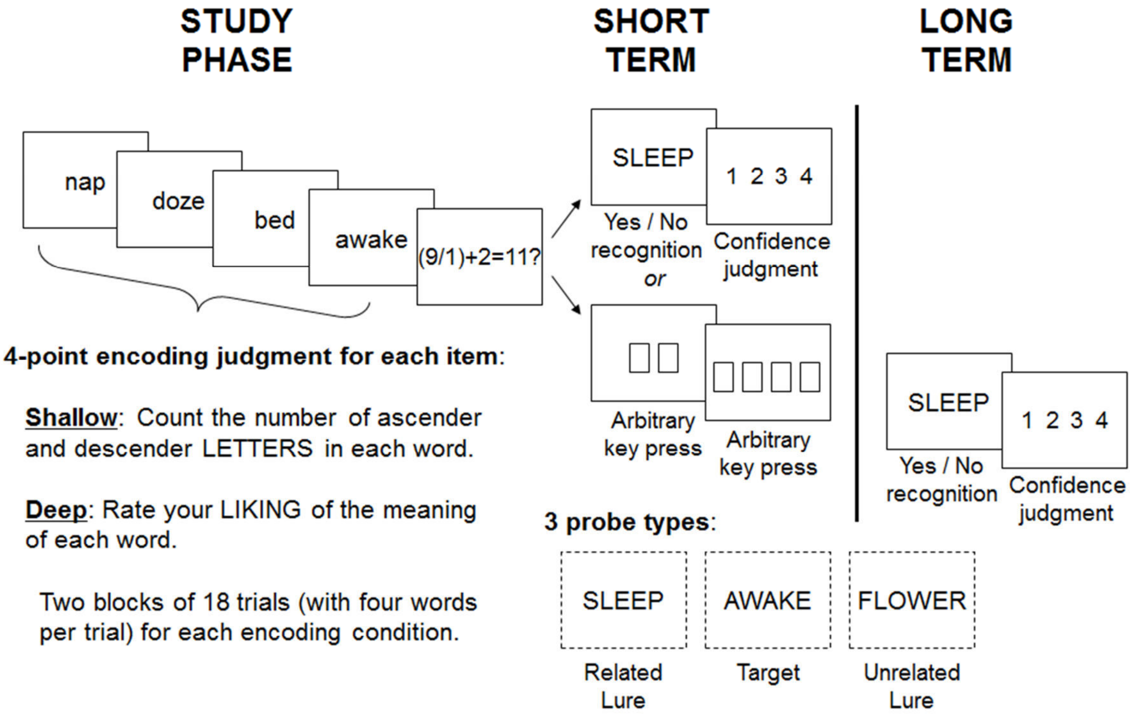


Figure 4.2 Experiment 1 proportion of “Yes” responses (error bars = SEM). In STM trials, the false memory effect (Related lure - Unrelated lure trials) does not differ between shallow and deep encoding conditions, while in LTM trials, the false memory effect is significant only for the deep encoding condition.

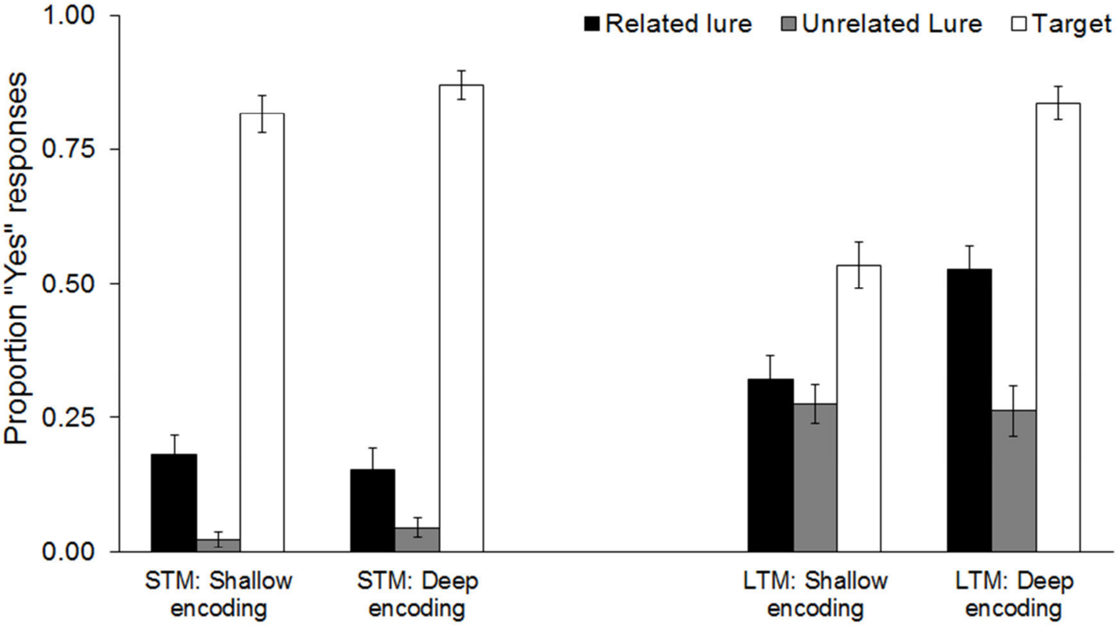


Figure 4.3 Experiment 1 mean confidence ratings by encoding condition, probe type, and delay (error bars = SEM; C = correct response).

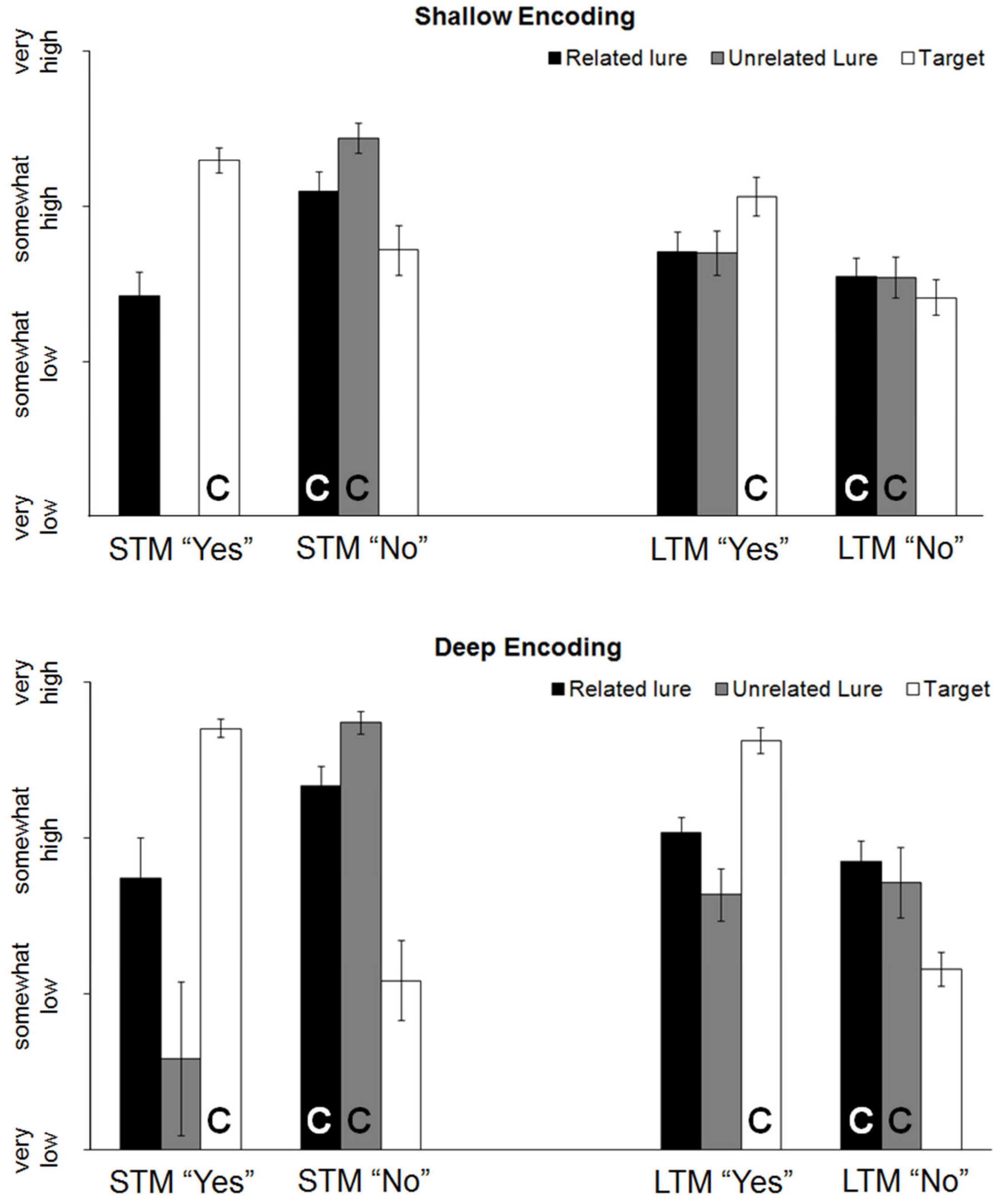


Table 4.1 Experiment 2 proportion of "Yes" responses.

A. Experiment 2A: Within-subjects manipulation

	Item-Specific Encoding				Relational Encoding			
	STM		LTM		STM		LTM	
	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Probe type								
Related lure	0.18	*0.04*	0.37	*0.04*	0.12	*0.03*	0.36	*0.04*
Unrelated lure	0.01	*0.01*	0.10	*0.03*	0.02	*0.01*	0.09	*0.03*
Target	0.91	*0.03*	0.69	*0.04*	0.91	*0.02*	0.69	*0.04*
Item-specific encoding first (n = 16)								
Probe type								
Related lure	0.21	*0.05*			0.09	*0.03*		
Unrelated lure	0.01	*0.01*			0.01	*0.01*		
Target	0.96	*0.03*			0.91	*0.03*		
Relational encoding first (n = 16)								
Probe type								
Related lure	0.16	*0.05*			0.14	*0.04*		
Unrelated lure	0.00	*0.00*			0.03	*0.02*		
Target	0.86	*0.05*			0.91	*0.03*		

B. Experiment 2B: Between-subjects manipulation

	Item-Specific Encoding (n = 14)				Relational Encoding (n = 14)			
	STM		LTM		STM		LTM	
	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Probe type								
Related lure	0.19	*0.06*	0.38	*0.06*	0.13	*0.03*	0.33	*0.06*
Unrelated lure	0.03	*0.01*	0.14	*0.04*	0.00	*0.00*	0.08	*0.02*
Target	0.92	*0.02*	0.76	*0.04*	0.94	*0.03*	0.69	*0.05*

Table 4.2 Experiment 2 proportion of “Remember”, “Know”, and “Guess” responses out of total proportion “Yes” responses.

A. Experiment 2A: Within-subjects manipulation

Item-Specific Encoding

	STM						LTM					
	“Remember”		“Know”		“Guess”		“Remember”		“Know”		“Guess”	
	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Probe type												
Related lure	0.21	*0.08*	0.25	*0.09*	0.54	*0.11*	0.24	*0.05*	0.30	*0.06*	0.47	*0.08*
Unrelated lure	0.00	*0.00*	0.00	*0.00*	1.00	*0.00*	0.03	*0.03*	0.33	*0.12*	0.64	*0.12*
Target	0.68	*0.05*	0.25	*0.04*	0.07	*0.02*	0.58	*0.05*	0.32	*0.05*	0.10	*0.03*

Relational Encoding

	STM						LTM					
	“Remember”		“Know”		“Guess”		“Remember”		“Know”		“Guess”	
	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	0.32	*0.12*	0.36	*0.13*	0.32	*0.12*	0.37	*0.07*	0.26	*0.07*	0.37	*0.08*
Unrelated lure	0.00	*0.00*	0.00	*0.00*	1.00	*0.00*	0.00	*0.00*	0.38	*0.14*	0.62	*0.14*
Target	0.78	*0.04*	0.14	*0.03*	0.08	*0.03*	0.62	*0.05*	0.24	*0.04*	0.14	*0.03*

B. Experiment 2B: Between-subjects manipulation

Item-Specific Encoding

	STM						LTM					
	“Remember”		“Know”		“Guess”		“Remember”		“Know”		“Guess”	
	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Probe type												
Related lure	0.22	*0.12*	0.21	*0.14*	0.57	*0.14*	0.18	*0.08*	0.39	*0.09*	0.43	*0.09*
Unrelated lure	0.00	*0.00*	0.00	*0.00*	1.00	*0.00*	0.10	*0.10*	0.20	*0.13*	0.70	*0.15*
Target	0.65	*0.07*	0.29	*0.06*	0.07	*0.02*	0.61	*0.06*	0.22	*0.04*	0.17	*0.04*

Relational Encoding

	STM						LTM					
	“Remember”		“Know”		“Guess”		“Remember”		“Know”		“Guess”	
	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	0.26	*0.12*	0.46	*0.13*	0.28	*0.13*	0.30	*0.10*	0.32	*0.10*	0.38	*0.10*
Unrelated lure	0.00	*0.00*	0.00	*0.00*	0.00	*0.00*	0.00	*0.00*	0.25	*0.16*	0.75	*0.16*
Target	0.80	*0.07*	0.15	*0.06*	0.04	*0.02*	0.60	*0.07*	0.21	*0.05*	0.19	*0.06*

CHAPTER 5

CONCLUSION

Summary of findings

The present dissertation investigated the mechanisms of accurate and inaccurate remembering, with a particular focus on the outcomes of processes engaged at encoding. Chapter 2 examined the susceptibility of memory to distortion at short and long delays, and Chapter 4 extended these findings by characterizing the effects of encoding strategy manipulations on increasing or decreasing the false memory rate. Chapter 3 explored the potential for modifying memory function in older adults, through training to promote the engagement of effortful memory processes at encoding. Findings from this series of experiments highlighted similarities between distortions in short-term memory (STM) and long-term memory (LTM), and the importance of individual differences in older adults' predispositions, preferences, and pre-existing abilities for identifying "effective" memory strategies.

In Chapter 2, false memories were directly compared under STM and LTM conditions, using a modified version of the converging associates or DRM paradigm (Deese, 1959; Roediger & McDermott, 1995), in which lists of four semantically-related words were probed individually after 3-4 seconds, or with continuous recognition after all lists were encoded. Traditionally, sensory-based coding is considered more prevalent than meaning-based coding in STM, whereas the converse characterizes LTM, leading to the expectation that semantic memory distortions should be less robust in a canonical STM task. However, in two experiments, corrected false recognition rates, confidence ratings, and Remember/Know judgments revealed parallel false memory effects across short-term and long-term testing. These results indicate that compelling false memory illusions can be rapidly instantiated, and originate from processes that are not specific to LTM tasks – consistent with accumulating evidence from other domains in favor of unitary models of memory (Jonides, et al., 2008; Ranganath & Blumenfeld, 2005).

In Chapter 3, enforced encoding strategies were investigated as an approach to remediating age-related declines in memory function. Older adult participants completed seven sessions of memory training with a verbal LTM task, under conditions that either a) mandated a semantic, associative encoding strategy (forming sentences that linked to-be-remembered words), b) attempted to suppress such encoding by mandating rote rehearsal, or c) encouraged effort towards encoding (by enforcing study times) but allowed participants to choose their own strategies. Study time was fixed at 14 seconds per word in all three conditions, an important difference from an earlier study (Bissig & Lustig, 2007) which used the same training task under open-ended, self-paced conditions. Compared to those results, the environmental support of long encoding times in the present study benefitted training-task performance in all three conditions.

Furthermore, pre-existing verbal skills predicted training-task success in the experimenter-prescribed associative encoding condition, whereas this was not the case for the participant-chosen condition. These results suggest that the effectiveness of mnemonic strategy training is constrained by participants' initial ability levels, and that environmental support itself (such as conditions that encourage sufficient time and effort at encoding) may prompt older adults to engage individualized strategies that will maximally benefit memory performance.

In Chapter 4, the influence of encoding strategies was examined in the STM/LTM false memory task. A levels-of-processing manipulation (Craik & Lockhart, 1972) revealed that deep encoding (relative to shallow encoding) preserved accurate memory across delay and increased false memory at LTM, with minimal effects at STM. The time-dependent effects of processing depth on false memory formation contrasted with relatively time-invariant measurements of phenomenological experience, as levels of confidence in false recognition were statistically equivalent across delay. A comparison of item-specific and relational encoding (Einstein & Hunt, 1980; Hunt & Einstein, 1981; Hunt & McDaniel, 1993) produced effects on false memory that were more difficult to interpret. Contrary to results from previous DRM studies (e.g., McCabe, et al., 2004), false memories were not reduced by item-specific processing. The encoding manipulation outcomes also did not differ between within-subjects and between-subjects designs. However, relatively time-invariant measurements of false memory phenomenology were again observed, as the normalized incidence of *remember* responses to related lures were stable from STM to LTM.

Further considerations and practical implications

The three studies presented in this dissertation each shed light on the memorial consequences of processes engaged at encoding. The novel experimental paradigm introduced in Chapter 2 implicated a common processing basis for false memories under STM and LTM conditions, as the quantity (the relative incidence of false recognition) and quality (the accompanying phenomenology, or subjective experience) of memory distortions that have been widely documented in LTM (Gallo, 2010) were found to be established within seconds of encoding. The experiments in Chapter 2 did not constrain encoding strategies, and thus the levels of processing that word lists received overall are likely to have varied considerably across participants and perhaps even across trials, within individual participants. This assumption is supported by self-reports of strategy use from the participants assigned to the “Strategy Choice” condition in Chapter 3, whose post-training questionnaire responses indicated that the processing strategies they spontaneously chose covered a wide range from “shallow” to “deep”. Therefore, when encoding instructions are not experimentally controlled, outcome measures will likely reflect the combined effects of varying levels of processing. Experiment 1 from Chapter 4 underscored the importance of this factor, in that time-invariant false memory effects, replicating those from Chapter 2, were observed when the data were collapsed across encoding condition. Only when lists that received shallow encoding and lists that received deep encoding were analyzed separately did time-dependent effects emerge: from short-term to long-term testing, the false memory effect decreased for lists studied under shallow processing instructions, but increased for lists studied under deep processing instructions. Consistent with this pattern, Experiment 2 from Chapter 4 examined two deep encoding conditions (item-specific and relational processing), and found an overall increase in the rate of false recognition from STM to LTM.

Chapters 3 and 4 each offer insights for the design of future training programs to boost accurate memories and to reduce false memories. In the cognitive aging literature, approaches to improving older adults' memory performance through strategy-based training have encountered difficulties in benefitting low-ability individuals (Verhaeghen & Marcoen, 1996; Yesavage, et al., 1988), and limited transfer to untrained tasks (Lustig, et al., 2009). In the present research, environmental support in the form of fixed encoding time was found to produce training-task improvements as large as those resulting from instruction with an experimenter-provided integrative encoding strategy. Furthermore, only in the integrative encoding condition (with its emphasis on inter-trial associations) was better training-task performance not correlated with greater resistance to false memory in a DRM task, indicating that enhanced relational processing may have increased not only accurate memory, but also susceptibility to false memory for unstudied but related material (cf. Hege & Dodson, 2004; McCabe, et al., 2004). These results suggest that more benefits and fewer costs accrue when environmental support is provided but participants are allowed to devise their own encoding strategies, in accordance with a small number of published studies which have associated self-generated strategies with superior memory performance in older adults (Derwinger, et al., 2005; Derwinger, et al., 2003; Hill, et al., 1990). Thus, the findings reported in Chapter 3 identified intact but inefficiently-used encoding processes as a promising target for training in older adults, allowing for memory strategies to be tailored to participants' individual preferences and pre-existing strengths.

In the DRM literature, item-specific processing has been explored as an approach to reducing false memory rates that are typically inflated by other forms of deep encoding, such as relational processing. Its demonstrated effectiveness has been cited by some researchers as support for memory-based explanations, which claim that item-specific processing strengthens verbatim memory traces at encoding (Arndt & Reder, 2003; McCabe, et al., 2004; Smith & Hunt, 1998), and by other researchers as support for decision-based explanations, which claim that item-specific processing allows for decision criteria to be strategically adjusted at retrieval (Dodson & Schacter, 2002b; McCabe & Smith, 2006; Schacter, et al., 2001). In the present research, an item-specific encoding strategy was *not* found to have a protective effect against false memory illusions, at either STM or LTM, and whether manipulated within-subjects or between-subjects. However, it is not clear whether the findings reported in Chapter 4 accurately represented the influence of item-specific and relational processing at short and long delays, or whether they were obscured by low levels of encoding instruction adherence and other potential confounds.

Closing remarks

The research presented in this dissertation was motivated by the goal of understanding the mechanisms of accurate and inaccurate remembering, with the hope of making applied as well as theoretical contributions to research on human memory. I conducted three studies to explore the susceptibility of memory to distortion, and its potential for modification. Results from my investigations of semantic memory distortions implicated common processes operating under STM and LTM conditions, which can be mediated by strategies implemented at encoding. A memory training study with older adults generated complementary findings regarding the effects of encoding strategy instructions on vulnerability to false memory, and demonstrated that supportive conditions for participant-chosen strategies are sufficient to diminish age and ability differences in training task performance. Additional research should be conducted to clarify the short-term and long-term effects of item-specific encoding, and to explore ways that this knowledge can be leveraged to improve memory accuracy through behavioral interventions.

Appendix A

Word Lists Used in False Memory Experiments
(Chapters 2 and 4)

Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list	Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list
ACT	perform	0.29	0.22	ADULT	mature	0.17	0.12
	portray	0.25			responsible	0.13	
	drama	0.20			grown	0.12	
	pretend	0.12			kid	0.05	
AGAIN	repeat	0.29	0.12	ALONE	isolated	0.60	0.33
	never	0.08			solo	0.46	
	twice	0.07			lonely	0.22	
	stop	0.05			one	0.03	
ANGER	rage	0.54	0.40	ANNOY	bother	0.30	0.24
	mad	0.39			aggravate	0.30	
	enrage	0.38			irritate	0.20	
	fury	0.31			disturb	0.14	
ANSWER	question	0.77	0.58	ARGUE	debate	0.47	0.25
	reply	0.75			disagree	0.43	
	response	0.49			complain	0.08	
	solution	0.29			agree	0.00	
ARMY	navy	0.54	0.35	ATOM	molecule	0.46	0.26
	soldier	0.29			nucleus	0.32	
	infantry	0.28			neutron	0.23	
	marines	0.28			proton	0.04	
BABY	crib	0.84	0.59	BACK	front	0.71	0.35
	infant	0.75			spine	0.54	
	diaper	0.59			behind	0.10	
	carriage	0.19			forward	0.04	
BAKE	broil	0.27	0.15	BARBECUE	cookout	0.42	0.22
	oven	0.21			grill	0.26	
	cook	0.08			ribs	0.16	
	cake	0.03			sauce	0.04	
BASEMENT	cellar	0.29	0.16	BASKET	wicker	0.30	0.16
	attic	0.20			picnic	0.15	
	downstairs	0.11			waste	0.09	
	underground	0.05			laundry	0.08	

Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list	Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list
BEAUTIFUL	gorgeous	0.18	0.12	BEE	hive	0.81	0.68
	lovely	0.18			bumble	0.79	
	pretty	0.10			sting	0.60	
	ugly	0.03			buzz	0.52	
BLACK	white	0.66	0.41	BLOOD	plasma	0.82	0.51
	gray	0.37			donor	0.52	
	brown	0.34			vein	0.38	
	coal	0.29			vampire	0.33	
BOOM	sonic	0.28	0.14	BORING	dull	0.46	0.23
	bang	0.16			interesting	0.20	
	explosion	0.11			lecture	0.16	
	bomb	0.01			exciting	0.10	
BOX	cardboard	0.56	0.28	BREAD	rye	0.79	0.52
	carton	0.37			loaf	0.55	
	storage	0.11			butter	0.36	
	container	0.07			toast	0.36	
BROOM	dustpan	0.47	0.30	BUILDING	structure	0.41	0.23
	sweep	0.41			blocks	0.25	
	mop	0.18			construction	0.19	
	witch	0.12			empire	0.05	
BUTTERFLY	cocoon	0.41	0.17	BUY	purchase	0.58	0.34
	moth	0.27			sell	0.54	
	insect	0.00			store	0.14	
	wing	0.00			spend	0.10	
CARPET	rug	0.47	0.14	CHAIN	link	0.40	0.17
	floor	0.07			whip	0.18	
	magic	0.01			necklace	0.08	
	red	0.00			bicycle	0.02	
CHAIR	table	0.76	0.62	CHAOS	havoc	0.31	0.14
	swivel	0.59			anarchy	0.14	
	rocking	0.593			hectic	0.10	
	recliner	0.55			confusion	0.01	
CHEESE	cheddar	0.92	0.45	CHURCH	cathedral	0.72	0.48
	swiss	0.67			steeple	0.66	
	cracker	0.14			temple	0.28	
	mouse	0.07			preacher	0.25	
CITY	metropolis	0.54	0.42	CLAM	chowder	0.75	0.29
	town	0.53			oyster	0.36	
	urban	0.36			shell	0.03	
	suburb	0.27			mussel	0.03	

Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list	Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list
COLD	hot	0.67	0.64	COPY	duplicate	0.65	0.29
	shiver	0.67			carbon	0.37	
	arctic	0.64			original	0.12	
	frigid	0.57			photo	0.02	
CORN	cob	0.88	0.47	CUP	saucer	0.53	0.33
	husk	0.64			measuring	0.40	
	flake	0.28			mug	0.27	
	field	0.06			goblet	0.12	
DANCE	ballet	0.43	0.22	DESTROY	demolish	0.54	0.33
	ballerina	0.25			ruin	0.37	
	song	0.12			annihilate	0.23	
	aerobics	0.07			create	0.19	
DIRT	soil	0.72	0.62	DOCTOR	physician	0.80	0.59
	filth	0.69			nurse	0.55	
	mud	0.60			stethoscope	0.52	
	grime	0.46			surgeon	0.48	
DRY	towel	0.29	0.15	EGGS	omelet	0.84	0.54
	desert	0.13			bacon	0.52	
	moist	0.09			dozen	0.41	
	thirst	0.08			scramble	0.37	
FAIL	flunk	0.62	0.28	FINISH	done	0.68	0.39
	pass	0.25			start	0.40	
	succeed	0.15			complete	0.37	
	try	0.09			end	0.12	
FISH	trout	0.91	0.50	FLAG	banner	0.69	0.32
	cod	0.58			checkered	0.25	
	scales	0.29			stripes	0.18	
	shrimp	0.21			pole	0.16	
FLOWER	tulip	0.78	0.71	FOG	mist	0.50	0.33
	petals	0.75			haze	0.38	
	daisy	0.69			smog	0.33	
	vase	0.60			unclear	0.11	
FOOT	toe	0.61	0.44	FOREVER	eternity	0.66	0.46
	inch	0.47			infinity	0.56	
	ankle	0.36			always	0.38	
	shoe	0.32			endless	0.25	
FRAGILE	delicate	0.29	0.18	FRIEND	pal	0.77	0.65
	breakable	0.26			buddy	0.69	
	frail	0.14			companion	0.64	
	glass	0.02			neighbor	0.51	

Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list	Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list
FRUIT	kiwi	0.71	0.45	FUNNY	hilarious	0.81	0.53
	citrus	0.43			comedian	0.55	
	pear	0.35			humor	0.45	
	berry	0.30			clown	0.32	
GAS	fuel	0.66	0.34	GHOST	ghoul	0.65	0.47
	petroleum	0.31			goblin	0.44	
	station	0.24			phantom	0.41	
	oil	0.14			spook	0.38	
GIRL	boy	0.70	0.27	GIVE	take	0.41	0.30
	dolls	0.20			generous	0.30	
	female	0.10			share	0.24	
	dress	0.06			charity	0.23	
HELP	assist	0.84	0.35	HIGH	low	0.78	0.28
	aid	0.43			elevate	0.17	
	emergency	0.08			tower	0.08	
	wanted	0.04			jump	0.07	
HOLD	grasp	0.53	0.29	HORSE	saddle	0.88	0.74
	grip	0.25			pony	0.75	
	keep	0.22			gallop	0.66	
	carry	0.16			colt	0.65	
JOB	occupation	0.68	0.55	KING	throne	0.759	0.59
	employment	0.61			queen	0.73	
	career	0.55			crown	0.471	
	task	0.37			reign	0.383	
LETTER	envelope	0.49	0.35	LIE	fib	0.82	0.45
	stamp	0.38			untruthful	0.38	
	mailbox	0.27			deception	0.35	
	mail	0.25			dishonest	0.24	
LION	roar	0.61	0.40	LONG	short	0.22	0.13
	tamer	0.49			distance	0.15	
	tiger	0.31			hair	0.10	
	mane	0.20			far	0.04	
LOSE	win	0.76	0.46	MAN	woman	0.60	0.31
	find	0.40			lady	0.37	
	gain	0.38			handsome	0.14	
	defeat	0.29			male	0.13	
MANY	several	0.59	0.31	MAP	atlas	0.53	0.18
	few	0.47			chart	0.12	
	much	0.12			direction	0.03	
	plenty	0.06			world	0.03	

Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list	Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list
MARRY	wed	0.39	0.26	MATH	arithmetic	0.76	0.69
	engage	0.39			calculus	0.71	
	single	0.20			algebra	0.65	
	hitch	0.04			equation	0.65	
MIX	blend	0.56	0.38	MORNING	early	0.26	0.19
	stir	0.46			dawn	0.21	
	combine	0.26			dew	0.16	
	integrate	0.25			evening	0.14	
MOUNTAIN	climber	0.60	0.40	MOVIE	cinema	0.79	0.54
	hill	0.43			film	0.54	
	climb	0.29			theater	0.44	
	molehill	0.26			popcorn	0.37	
MUSCLE	flex	0.46		NEEDLE	thread	0.76	0.51
	weights	0.15			syringe	0.52	
	strength	0.12			haystack	0.42	
	tone	0.03			injection	0.33	
NUT	cashew	0.75	0.47	PAN	skillet	0.48	0.22
	pecan	0.47			pot	0.29	
	almond	0.34			fry	0.07	
	squirrel	0.30			dish	0.04	
PARK	lot	0.30	0.16	PEN	quill	0.64	0.37
	bench	0.12			pencil	0.48	
	recreation	0.11			marker	0.26	
	playground	0.10			write	0.13	
PIANO	keyboard	0.35	0.19	PIG	hog	0.74	0.44
	organ	0.27			pork	0.59	
	guitar	0.08			sow	0.29	
	ivory	0.06			sty	0.12	
PRESENT	gift	0.61	0.35	PULL	tug	0.58	0.31
	past	0.43			push	0.40	
	absent	0.29			drag	0.18	
	future	0.06			stretch	0.07	
RAIN	umbrella	0.70	0.38	RENT	own	0.35	0.14
	storm	0.37			lease	0.13	
	hail	0.27			apartment	0.04	
	puddle	0.18			monthly	0.03	
RING	diamond	0.63	0.31	RIVER	creek	0.40	0.30
	bell	0.40			stream	0.32	
	jewelry	0.12			flow	0.28	
	phone	0.08			bridge	0.20	

Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list	Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list
ROCK	boulder	0.66	0.39	ROOF	shingle	0.61	0.27
	stone	0.63			ceiling	0.35	
	solid	0.17			tar	0.09	
	roll	0.08			tin	0.03	
ROPE	knot	0.39	0.22	ROUGH	sandpaper	0.43	0.33
	string	0.23			smooth	0.42	
	noose	0.15			coarse	0.29	
	twine	0.12			tough	0.19	
SAFE	vault	0.25	0.14	SHIRT	blouse	0.65	0.40
	secure	0.23			sleeves	0.35	
	guard	0.04			collar	0.34	
	lock	0.02			shorts	0.25	
SHOVEL	dig	0.32	0.17	SHY	bashful	0.73	0.42
	spade	0.16			timid	0.69	
	pail	0.10			introvert	0.18	
	rake	0.09			outgoing	0.08	
SINK	drain	0.32	0.21	SLEEP	nap	0.73	0.67
	float	0.25			doze	0.68	
	faucet	0.16			bed	0.64	
	bathroom	0.09			awake	0.62	
SLOW	fast	0.60	0.45	SMART	intelligent	0.71	0.47
	snail	0.49			genius	0.43	
	turtle	0.37			wise	0.43	
	sluggish	0.34			knowledge	0.29	
SMELL	aroma	0.68	0.61	SMOKE	cigar	0.51	0.43
	scent	0.63			cigarette	0.45	
	whiff	0.58			pipe	0.42	
	stench	0.56			tobacco	0.34	
SNAKE	slither	0.63	0.54	SNEEZE	allergy	0.41	0.22
	serpent	0.57			cough	0.21	
	reptile	0.51			handkerchief	0.15	
	rattle	0.46			tissue	0.09	
SOAP	lather	0.67	0.34	SOFT	hard	0.56	0.37
	suds	0.60			loud	0.33	
	shower	0.05			tender	0.30	
	bubble	0.04			fluffy	0.27	
SORRY	apology	0.58	0.23	SPICE	oregano	0.49	0.29
	regret	0.21			herb	0.28	
	guilt	0.06			cinnamon	0.20	
	forgive	0.05			seasoning	0.20	

Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list	Related Lure	Studied Associates	BAS to Related Lure	Mean BAS for list
SPIDER	web	0.85	0.59	SQUARE	circle	0.63	0.37
	tarantula	0.74			triangle	0.43	
	arachnid	0.70			round	0.39	
	creepy	0.06			shape	0.03	
STOMACH	abdomen	0.57	0.38	STRESS	tension	0.48	0.19
	belly	0.40			pressure	0.14	
	intestine	0.31			anxiety	0.08	
	ulcer	0.23			strain	0.07	
SWEET	honey	0.45	0.43	TEETH	gums	0.70	0.40
	bitter	0.44			braces	0.65	
	sugar	0.43			mouth	0.19	
	sour	0.41			tongue	0.07	
THIEF	crook	0.46	0.31	TRASH	garbage	0.46	0.33
	robber	0.36			rubbish	0.40	
	burglar	0.26			debris	0.27	
	bandit	0.17			dump	0.22	
TRIP	journey	0.46	0.25	VOTE	ballot	0.52	0.24
	vacation	0.42			election	0.30	
	travel	0.06			register	0.09	
	baggage	0.05			campaign	0.03	
WHOLE	half	0.45	0.16	WINDOW	pane	0.83	0.55
	part	0.14			sill	0.68	
	piece	0.03			shutter	0.48	
	all	0.02			curtain	0.19	

Note: BAS = Backward Associative Strength

Grouping of Lists for Experiments in Chapter 2

A1	A2	B1	B2	C1	C2	D1	D2
HORSE	CHAIR	FLOWER	COLD	BEE	SLEEP	MATH	FRIEND
JOB	DIRT	ANSWER	SMELL	KING	BABY	DOCTOR	SPIDER
NEEDLE	WINDOW	SNAKE	BLOOD	EGGS	BREAD	MOVIE	FUNNY
FISH	LOSE	FOREVER	CHURCH	NUT	GHOST	SMART	CORN
SMOKE	CHEESE	SLOW	SWEET	LIE	PIG	FRUIT	FOOT
TEETH	CITY	SHIRT	SHY	MOUNTAIN	BLACK	LION	ANGER
MIX	SOFT	RAIN	SQUARE	FINISH	PEN	ROCK	STOMACH
GAS	BACK	SOAP	ARMY	PRESENT	HELP	BUY	LETTER
TRASH	THIEF	ROUGH	FLAG	DESTROY	CUP	FOG	ALONE
BROOM	MAN	CLAM	MANY	GIVE	PULL	RIVER	RING
SPICE	FAIL	HOLD	ROOF	HIGH	COPY	BOX	GIRL
VOTE	ATOM	BORING	MARRY	BUILDING	TRIP	ANNOY	ARGUE
ROPE	SINK	BARBECUE	ACT	SORRY	DANCE	PAN	SNEEZE
BUTTERFLY	MORNING	CHAIN	MUSCLE	MAP	STRESS	FRAGILE	PIANO
SHOVEL	DRY	BAKE	WHOLE	BASKET	PARK	BOOM	BASEMENT
ADULT	SAFE	BEAUTIFUL	CARPET	CHAOS	AGAIN	LONG	RENT

Note: Mean BAS for each group of 16 lists = 0.35

Grouping of Lists for Experiments in Chapter 4

A1	A2	B1	B2	C1	C2	D1	D2
JOB	HORSE	SHIRT	SLOW	KING	PIG	DOCTOR	MATH
FISH	NEEDLE	SOAP	RAIN	NUT	PEN	SMART	MOVIE
BUTTERFLY	ROPE	CLAM	ROUGH	MAP	CUP	FRAGILE	PAN
TEETH	SMOKE	ANSWER	FLOWER	MOUNTAIN	SLEEP	LION	FRUIT
GAS	MIX	FOREVER	SNAKE	PRESENT	BREAD	BUY	ROCK
BROOM	TRASH	CHAIN	BARBECUE	GIVE	DANCE	RIVER	FOG
CHAIR	DIRT	COLD	SMELL	BEE	BLACK	FRIEND	SPIDER
WINDOW	LOSE	BLOOD	CHURCH	EGGS	HELP	FUNNY	CORN
SINK	MORNING	ACT	MUSCLE	SORRY	PULL	SNEEZE	PIANO
CHEESE	CITY	SWEET	SHY	LIE	BABY	FOOT	ANGER
SOFT	BACK	SQUARE	ARMY	FINISH	GHOST	STOMACH	LETTER
THIEF	MAN	FLAG	MANY	DESTROY	STRESS	ALONE	RING

Note: Mean BAS for each group of 6 lists = 0.4

Appendix B

Response Time Data from False Memory Experiments (Chapters 2 and 4)

Table B.1 Chapter 2 median response times (in ms) by probe type.

Experiment 1

	STM				LTM			
	"Yes"		"No"		"Yes"		"No"	
Probe type	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	1781.4	*124.6*	**1539.5**	***67.7***	1575.8	*63.6*	**1687.4**	***83.0***
Unrelated lure	1451.0	*434.9*	**1236.6**	***57.1***	1606.6	*73.8*	**1596.2**	***58.5***
Target	**1251.6**	***61.5***	1660.6	*146.6*	**1433.2**	***51.6***	1694.7	*76.3*
Semantic Interference Effect			302.9	*51.4*			91.2	*64.8*

Experiment 2

	STM				LTM			
	"Yes"		"No"		"Yes"		"No"	
Probe type	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	1408.8	*82.1*	**1435.4**	***52.0***	1468.0	*59.8*	**1446.3**	***47.4***
Unrelated lure	1223.8	*183.6*	**1164.5**	***44.6***	1466.1	*59.1*	**1443.0**	***53.1***
Target	**1172.4**	***43.3***	1538.7	*140.7*	**1324.2**	***52.0***	1491.5	*71.7*
Semantic Interference Effect			288.6	*45.2*			3.3	*51.0*

Note: Correct responses are in **bold**. Semantic Interference Effect = average correct response time on related lure trials - average correct response time on unrelated lure trials.

Table B.2 Chapter 4, Experiment 1 median response times (in ms) by probe type.

Shallow Encoding

	STM				LTM			
	"Yes"		"No"		"Yes"		"No"	
Probe type	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	1754.1	*108.7*	**1612.8**	***61.2***	1487.0	*89.3*	**1535.0**	***70.1***
Unrelated lure	1864.8	*249.9*	**1257.0**	***53.4***	1462.9	*76.0*	**1463.1**	***67.5***
Target	**1432.4**	***50.5***	1793.5	*102.7*	**1374.3**	***52.2***	1594.5	*85.1*
SIE			355.8	*71.5*			71.0	*61.9*

Deep Encoding

	STM				LTM			
	"Yes"		"No"		"Yes"		"No"	
Probe type	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	1680.0	*158.5*	**1558.7**	***66.5***	1424.4	*67.7*	**1670.5**	***70.4***
Unrelated lure	1364.8	*394.6*	**1292.6**	***55.1***	1520.1	*96.5*	**1410.8**	***57.7***
Target	**1290.4**	***37.0***	1834.3	*189.9*	**1268.0**	***52.8***	1520.0	*108.8*
SIE			266.1	*94.3*			264.6	*86.5*

Note: Correct responses are in **bold**. Semantic Interference Effect (SIE) = average correct response time on related lure trials - average correct response time on unrelated lure trials.

Table B.3 Chapter 4, Experiment 2 median response times (in ms) by probe type.

A. Experiment 2A: Within-subjects manipulation

Item-Specific Encoding

	STM				LTM			
	"Yes"		"No"		"Yes"		"No"	
Probe type	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	1584.3	*134.4*	1532.6	*85.8*	1663.3	***101.4***	**1572.6**	***90.8***
Unrelated lure	914.0	*0.0*	1144.4	*43.3*	1655.1	***145.6***	**1442.6**	***69.4***
Target	**1270.2**	***45.3***	1348.1	*94.3*	1375.3	*52.4*	1542.2	*85.8*
SIE			388.2	*61.0*			130.0	*60.0*

Relational Encoding

	STM				LTM			
	"Yes"		"No"		"Yes"		"No"	
Probe type	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	1909.7	*141.6*	1510.7	*76.7*	1569.1	*68.4*	**1652.8**	***102.1***
Unrelated lure	1092.7	*642.3*	1147.9	*46.5*	1453.3	*170.0*	**1423.6**	***71.0***
Target	**1300.8**	***53.8***	1674.7	*254.8*	1358.0	*52.5*	1635.6	*81.5*
SIE			362.8	*63.4*			229.2	*76.0*

B. Experiment 2B: Between-subjects manipulation

Item-Specific Encoding

	STM				LTM			
	"Yes"		"No"		"Yes"		"No"	
Probe type	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	1850.3	*100.9*	**1551.5**	***94.8***	1726.8	*103.3*	**1498.8**	***91.7***
Unrelated lure	1614.3	*605.0*	**1222.4**	***77.3***	1736.9	*106.0*	**1349.7**	***63.4***
Target	**1306.8**	***47.6***	2064.6	*149.9*	**1405.4**	***87.2***	1625.1	*116.3*
SIE			329.2	*87.3*			149.1	*86.9*

Relational Encoding

	STM				LTM			
	"Yes"		"No"		"Yes"		"No"	
Probe type	M	*SEM*	M	*SEM*	M	*SEM*	M	*SEM*
Related lure	1724.0	*144.4*	**1443.8**	***67.2***	1692.0	*105.9*	**1603.8**	***136.5***
Unrelated lure	--	--	**1163.6**	***76.4***	2041.3	*194.3*	**1397.8**	***103.0***
Target	**1208.7**	***58.4***	1680.5	*295.9*	**1407.3**	***70.7***	1580.5	*140.7*
SIE			280.2	*66.2*			206.0	*92.3*

Note: Correct responses are in **bold**. Semantic Interference Effect (SIE) = average correct response time on related lure trials - average correct response time on unrelated lure trials.

References

Anastasi, J. S., Rhodes, M. G., & Burns, M. C. (2000). Distinguishing between memory illusions and actual memories using phenomenological measurements and explicit warnings. *American Journal of Psychology, 113*(1), 1-26.

Armitage, S. G. (1946). An analysis of certain psychological tests used for the evaluation of brain injury. *Psychological Monographs, 60*(1), 1-48.

Arndt, J. (2006). Distinctive information and false recognition: The contribution of encoding and retrieval factors. *Journal of Memory and Language, 54*(1), 113-130.

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language, 48*(1), 1-15.

Atkins, A. S., & Reuter-Lorenz, P. A. (2008). False working memories? Semantic distortion in a mere 4 seconds. *Memory & Cognition, 36*(1), 74-81.

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53*(6), 452-471.

Baddeley, A. D. (1966). Influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology, 18*, 302-309.

Baddeley, A. D. (1986). *Working Memory*. Oxford, UK: Clarendon Press.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., Willis, S. L. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *JAMA: Journal of the American Medical Association, 288*(18), 2271-2281.

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K. B., & Yerys, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology, 16*(3-5), 361-384.

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

Baltes, P. B., Sowarka, D., & Kliegl, R. (1989). Cognitive training research on fluid intelligence in old-age: What can older adults achieve by themselves? *Psychology and Aging, 4*(2), 217-221.

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge, UK: Cambridge University Press.

Berry, J. M., West, R. L., & Dennehey, D. M. (1989). Reliability and validity of the memory self-efficacy questionnaire. *Developmental Psychology, 25*(5), 701-713.

Bissig, D., & Lustig, C. (2007). Who benefits from memory training? *Psychological Science, 18*(8), 720-726.

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

Brainerd, C. J., & Reyna, V. F. (2005). *The Science of False Memory*. New York: Oxford University Press.

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in "associative" false memory. *Psychonomic Bulletin & Review, 15*(6), 1035-1053.

Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., Reed, B. R. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General, 130*(4), 746-763.

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake & J. N. Towse (Eds.), *Variation in working memory* (pp. 76–106). Oxford University Press.

Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and Aging, 3*(3), 249-257.

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*(2), 277-287.

Butler, K. M., McDaniel, M. A., McCabe, D. P., & Dornburg, C. C. (2010). The influence of distinctive processing manipulations on older adults' false memory. *Aging, Neuropsychology, and Cognition, 17*(2), 129-159.

Calero, M. D., & Navarro, E. (2007). Cognitive plasticity as a modulating variable on the effects of memory training in elderly persons. *Archives of Clinical Neuropsychology, 22*(1), 63-72.

Cerella, J. (1985). Information-processing rates in the elderly. *Psychological Bulletin, 98*(1), 67-83.

Coane, J. H., & McBride, D. M. (2006). The role of test structure in creating false memories. *Memory & Cognition, 34*(5), 1026-1036.

Coane, J. H., McBride, D. M., Raulerson, B. A., & Jordan, J. S. (2007). False memory in a short-term memory task. *Experimental Psychology, 54*(1), 62-70.

Colbert, J. M., & McBride, D. M. (2007). Comparing decay rates for accurate and false memories in the DRM paradigm. *Memory & Cognition, 35*(7), 1600-1609.

Cotel, S. C., Gallo, D. A., & Seamon, J. G. (2008). Evidence that nonconscious processes are sufficient to produce false memories. *Consciousness and Cognition, 17*(1), 210-218.

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

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Craik, F. I. M., & Rabinowitz, J. C. (1985). The effects of presentation rate and encoding task on age-related memory deficits. *Journal of Gerontology, 40*(3), 309-315.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294.

D'Esposito, M., Postle, B. R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Experimental Brain Research, 133*(1), 3-11.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17-22.

Dehon, H., & Bredart, 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*(1), 191-197.

Dennis, N. A., Kim, H., & Cabeza, R. (2007). Effects of aging on true and false memory formation: An fMRI study. *Neuropsychologia, 45*(14), 3157-3166.

Derwinger, A., Stigsdotter Neely, A., & Bäckman, L. (2005). Design your own memory strategies! Self-generated strategy training versus mnemonic training in old age: An 8-month follow-up. *Neuropsychological Rehabilitation, 15*(1), 37-54.

Derwinger, A., Stigsdotter Neely, A., Persson, M., Hill, R. D., & Bäckman, L. (2003). Remembering numbers in old age: Mnemonic training versus self-generated strategy training. *Aging, Neuropsychology, and Cognition, 10*(3), 202-214.

Dodson, C. S., Koutstaal, W., & Schacter, D. L. (2000). Escape from illusion: reducing false memories. *Trends in Cognitive Sciences, 4*(10), 391-397.

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8*(1), 155-161.

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

Dodson, C. S., & Schacter, D. L. (2002b). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language, 46*(4), 782-803.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*(2), 170-177.

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

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13*(4), 597-607.

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition, 29*(2), 247-253.

Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology, 41*(2), 389-400.

Dunlosky, J., Kubat-Silman, A. K., & Hertzog, C. (2003). Training monitoring skills improves older adults' self-paced associative learning. *Psychology and Aging, 18*(2), 340-345.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 588-598.

Flegal, K. E., Atkins, A. S., & Reuter-Lorenz, P. A. (2010). False memories seconds later: The rapid and compelling onset of illusory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1331-1338.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 120-128.

Gallo, D. A. (2006). *False Memory Research in DRM and Related Tasks*. New York: Psychology Press.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38*(7), 833-848.

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*(2), 271-276.

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: evidence for associative activation and monitoring. *Journal of Memory and Language, 47*(3), 469-497.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*(4), 309-313.

Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition, 19*(6), 617-623.

Geraci, L., & McCabe, D. P. (2006). Examining the basis for illusory recollection: The role of remember/know instructions. *Psychonomic Bulletin & Review, 13*(3), 466-473.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 351-360.

Glisky, E. L., Rubin, S. R., & Davidson, P. S. R. (2001). Source memory in older adults: An encoding or retrieval problem? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1131-1146.

Hashtroudi, S., Chrosniak, L. D., & Johnson, M. K. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and Aging, 5*(1), 119-126.

Hege, A. C. G., & Dodson, C. S. (2004). Why distinctive information reduces false memories: Evidence for both impoverished relational-encoding and distinctiveness heuristic accounts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(4), 787-795.

Hertzog, C., Dunlosky, J., & Robinson, A. E. (unpublished manuscript). *Intellectual Abilities and Metacognitive Beliefs Influence Spontaneous Use of Effective Encoding Strategies*.

Hertzog, C., Price, J., & Dunlosky, J. (in press). Age differences in the effects of experimenter-instructed versus self-generated strategy use. *Experimental Aging Research*.

Hill, R. D., Allen, C., & Gregory, K. (1990). Self-generated mnemonics for enhancing free-recall performance in older learners. *Experimental Aging Research, 16*(3), 141-145.

Hill, R. D., Yesavage, J. A., Sheikh, J., & Friedman, L. (1989). Mental status as a predictor of response to memory training in older adults. *Educational Gerontology, 15*(6), 633-639.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*(5), 497-514.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*(4), 421-445.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review, 4*(4), 577-581.

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 3-22.

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

Jennings, J. M., & Jacoby, L. L. (2003). Improving memory in older adults: Training recollection. *Neuropsychological Rehabilitation, 13*(4), 417-440.

Jennings, J. M., Webster, L. M., Kleykamp, B. A., & Dagenbach, D. (2005). Recollection training and transfer effects in older adults: Successful use of a repetition-lag procedure. *Aging, Neuropsychology, and Cognition, 12*(3), 278-298.

Jones, S., Nyberg, L., Sandblom, J., Stigsdotter Neely, A., Ingvar, M., Petersson, K. M., & Bäckman, L. (2006). Cognitive and neural plasticity in aging: General and task-specific limitations. *Neuroscience and Biobehavioral Reviews, 30*(6), 864-871.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*(1), 193-224.

Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., & Schimmel, H. (1983). Validation of a short orientation-memory-concentration test of cognitive impairment. *American Journal of Psychiatry, 140*(6), 734-739.

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

Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of ageing. *Cognitive Neuropsychology, 16*(3-5), 399-415.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review, 114*(4), 954-993.

Kliegl, R., Smith, J., & Baltes, P. B. (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology, 25*(2), 247.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37*(4), 555-583.

Koutstaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. A. (1999). Reducing gist-based false recognition in older adults: Encoding and retrieval manipulations. *Psychology and Aging, 14*(2), 220-237.

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*(5), 954-963.

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1998). Memory illusions and consciousness: Examining the phenomenology of true and false memories. *Current Psychology, 16*(3-4), 181-224.

Lampinen, J. M., Ryals, D. B., & Smith, K. (2008). Compelling untruths: The effect of retention interval on content borrowing and vivid false memories. *Memory, 16*(2), 149-156.

Lampinen, J. M., & Schwartz, R. M. (2000). The impersistence of false memory persistence. *Memory, 8*(6), 393-400.

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*(1), 19-31.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: Example of interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*(5), 585-589.

Logan, J. M., Sanders, A. L., Snyder, A. Z., Morris, J. C., & Buckner, R. L. (2002). Under-recruitment and nonselective recruitment: Dissociable neural mechanisms associated with aging. *Neuron, 33*(5), 827-840.

Lustig, C., & Flegal, K. E. (2008). Targeting latent function: Encouraging effective encoding for successful memory training and transfer. *Psychology and Aging, 23*(4), 754-764.

Lustig, C., Shah, P., Seidler, R., & Reuter-Lorenz, P. A. (2009). Aging, training, and the brain: A review and future directions. *Neuropsychology Review, 19*(4), 504-522.

Luszcz, M. A., Roberts, T. H., & Mattiske, J. (1990). Use of relational and item-specific information in remembering by older and younger adults. *Psychology and Aging, 5*(2), 242-249.

McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychonomic Bulletin & Review, 11*(6), 1074-1079.

McCabe, D. P., & Smith, A. D. (2006). The distinctiveness heuristic in false recognition and false recall. *Memory, 14*(5), 570-583.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35*(2), 212-230.

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

McDermott, K. B., & Roediger, H. L. (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*(3), 508-520.

Murphy, M. D., Sanders, R. E., Gabriesheski, A. S., & Schmitt, F. A. (1981). Metamemory in the aged. *Journals of Gerontology, 36*(2), 185-193.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1170-1187.

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging, 22*(1), 202-208.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*, from http://www.usf.edu/FreeAssociation/

Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory, 9*(1), 53-71.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*(6), 838-848.

Park, D. C., Gutchess, A. H., Meade, M. L., & Stine-Morrow, E. A. L. (2007). Improving cognitive function in older adults: Nontraditional approaches. *Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 62*, 45-52.

Paxton, J. L., Barch, D. M., Racine, C. A., & Braver, T. S. (2008). Cognitive control, goal maintenance, and prefrontal function in healthy aging. *Cerebral Cortex, 18*(5), 1010-1028.

Paxton, J. L., Barch, D. M., Storandt, M., & Braver, T. S. (2006). Effects of environmental support and strategy training on older adults' use of context. *Psychology and Aging, 21*(3), 499-509.

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*(2), 261-285.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*(1), 89-102.

Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences, 9*(8), 374-380.

Rebok, G. W., & Balcerak, L. J. (1989). Memory self-efficacy and performance differences in young and old adults: The effect of mnemonic training. *Developmental Psychology, 25*(5), 714-721.

Rebok, G. W., Carlson, M. C., & Langbaurn, J. B. S. (2007). Training and maintaining memory abilities in healthy older adults: Traditional and novel approaches. *Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 62*, 53-61.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*(1), 1-75.

Rhodes, M. G., & Anastasi, J. S. (2000). The effects of a levels-of-processing manipulation on false recall. *Psychonomic Bulletin & Review, 7*(1), 158-162.

Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*(3), 231-237.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803-814.

Roediger, H. L., McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in producing false memories. In M. Conway, S. Gathercole & C. Cornoldi (Eds.), *Theories of memory* (Vol. 2, pp. 187-245). Hove, England: Psychology Press.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*(3), 385-405.

Rose, N. S., Myerson, J., Roediger, H. L., & Hale, S. (2010). Similarities and differences between working memory and long-term memory: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2), 471-483.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403-428.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age-differences in working memory. *Developmental Psychology, 27*(5), 763-776.

Schacter, D. L., Cendan, D. L., Dodson, C. S., & Clifford, E. R. (2001). Retrieval conditions and false recognition: Testing the distinctiveness heuristic. *Psychonomic Bulletin & Review, 8*(4), 827-833.

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), 1-24.

Schacter, D. L., Kaszniak, A. W., Kihlstrom, J. F., & Valdiserri, M. (1991). The relation between source memory and aging. *Psychology and Aging, 6*(4), 559-568.

Schacter, D. L., Koutstaal, W., & Norman, K. A. (1997). False memories and aging. *Trends in Cognitive Sciences, 1*(6), 229-236.

Schacter, D. L., & Wagner, A. D. (1999). Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus, 9*(1), 7-24.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20*(1), 11-21.

Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., & Schwartz, M. A. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory & Cognition, 30*(7), 1054-1064.

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review, 5*(4), 710-715.

Sunderland, A., Harris, J. E., & Baddeley, A. D. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 341-357.

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition, 29*(3), 424-432.

Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language, 52*(1), 71-86.

Thompson, L. A., & Kliegl, R. (1991). Adult age effects of plausibility on memory: The role of time constraints during encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(3), 542-555.

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7*(2), 233-256.

Touron, D. R., & Hertzog, C. (2004). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging, 19*(3), 452-466.

Tse, C. S., & Neely, J. H. (2007). Semantic and repetition priming effects for Deese/Roediger-McDermott (DRM) critical items and associates produced by DRM and unrelated study lists. *Memory & Cognition, 35*(5), 1047-1066.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*(1), 1-12.

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*(2), 230-241.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

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

Verhaeghen, P., & Marcoen, A. (1994). Production deficiency hypothesis revisited: Adult age-differences in strategy use as a function of processing resources. *Aging and Cognition, 1*(4), 323-338.

Verhaeghen, P., & Marcoen, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and Aging, 11*(1), 164-178.

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging, 7*(2), 242-251.

Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory verbal short-term memory. *Brain, 92*(4), 885-896.

Wood, L. E., & Pratt, J. D. (1987). Pegword mnemonic as an aid to memory in the elderly: A comparison of 4 age-groups. *Educational Gerontology, 13*(4), 325-339.

Yesavage, J. A., & Rose, T. L. (1984). The effects of a face-name mnemonic in young, middle-aged, and elderly adults. *Experimental Aging Research, 10*(1), 55-57.

Yesavage, J. A., Sheikh, J., Tanke, E. D., & Hill, R. (1988). Response to memory training and individual-differences in verbal intelligence and state anxiety. *American Journal of Psychiatry, 145*(5), 636-639.

Yesavage, J. A., Sheikh, J. I., Friedman, L., & Tanke, E. (1990). Learning mnemonics: Roles of aging and subtle cognitive impairment. *Psychology and Aging, 5*(1), 133-137.

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