EXAMINING AGE-RELATED DIFFERENCES IN KNOWLEDGE UPDATING IN A CATEGORIZED LIST-LEARNING TASK

A Dissertation
Presented to
The Academic Faculty

By

Jarrod Charles Hines

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Psychology

Georgia Institute of Technology

May, 2013

EXAMINING AGE-RELATED DIFFERENCES IN KNOWLEDGE UPDATING
IN A CATEGORIZED LIST-LEARNING TASK

Approved by:

Dr. Christopher Hertzog, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Audrey Duarte
School of Psychology
Georgia Institute of Technology

Dr. Ute Fischer
School of Literature, Communication,
And Culture
Georgia Institute of Technology

Dr. Wendy Rogers
School of Psychology
Georgia Institute of Technology

Dr. Paul Verhaeghen
School of Psychology
Georgia Institute of Technology

Date Approved: March 15, 2013

ACKNOWLEDGEMENTS

I would like to give special thanks to Ron and Lynn Hines for their lifelong universal support, Chris Hertzog for his seemingly infinite patience and wisdom, and Hi Shin Shim for helping me maintain my happiness and sanity throughout my time at Georgia Tech.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CJ	Confidence Judgment
JOL	Judgment of Learning
RT	Response Time

SUMMARY

Distinctive encoding is the processing of unique item-specific information in the context of more general relational or organizational information. It enhances memory performance for both younger and older adults (Smith, 2006). The current work examined how adults use distinctive encoding to aid their free recall performance and whether task experience alters subsequent use of a distinctive encoding strategy. At study participants saw a series of five-item taxonomically categorized lists (e.g., FRUITS). They were first required to generate a category-consistent label (e.g., TASTY FRUIT). In the guided condition, they were then required to generate a single word representing either (1) another category-consistent characteristic (e.g., GROWS) or (2) a characteristic that distinguished a study target from the other items (e.g., FUZZY for the target KIWI). In the self-initiated condition, participants were allowed to select an encoding strategy on their own. After test, all participants completed a second study-test phase with self-initiated strategies. Younger adults initially rated distinctive encoding as more effective, relative to relational encoding, than did older adults, and this difference persisted after test experience, indicating an age difference in learning about the relative superiority of distinctive processing. Consistent with these ratings, distinctive encoding was implemented more so by unguided younger adults than older adults in phase 1. However, both strategy use and recall performance were similar across age and study conditions in phase 2. Both older and younger adults were capable of utilizing distinctive encoding effectively following task experience, although perceptions of strategic effectiveness did not always correspond to self-initiated study behaviors.

INTRODUCTION

Processing unique item-specific information in the context of more general relational or organizational information is referred to as distinctive encoding, and this has been shown to enhance memory performance in a variety of tasks (see Smith, 2006, for a review). The current work examined the degree to which younger and older adults were able to (1) appreciate the memory benefit afforded by the use of an encoding strategy that highlights the distinctive properties of to-be-remembered information and (2) make use of this strategy effectively during an episodic memory task. The current work also assessed the degree to which younger and older adults were capable of updating their knowledge of the relative effectiveness of distinctive and purely relational encoding strategies during an initial study-test phase and use this information to increase their quality of learning during a second phase.

1.1 Distinctiveness Effects in Episodic Memory

The von Restorff effect (von Restorff, 1933, as cited by Hunt, 1995), is one of the earliest examples of the benefit of so-called distinctive processing on memory: a unique, or distinct, item in a list (e.g., the isolated trigram "RFN" among an otherwise homogenous list of numbers) is generally remembered better than the other list members. Subsequent research explained this isolation effect and similar memory phenomena with theoretical accounts of encoding and retrieval processes. With respect to encoding processes, the general idea is that distinct stimuli benefit from greater rehearsal or semantic elaboration of the item and the context in which it appears during study (e.g., Hirshman, Whelley, & Palij, 1989; Schmidt, 1991) or from a more thorough evaluation of their physical characteristics (e.g., the spelling or pronunciation of orthographically uncommon words, Geraci & Rajaram, 2002). The unique characteristics of encoded items allow them to stand apart (i.e., be distinct) from other items that exist in memory during a retrieval process.

During retrieval, the activation of item characteristics that are either shared with other studied items or unique to only the target item allows one to distinguish a specific target item from other items that are available in memory (e.g., Hunt & McDaniel, 1993). Shared characteristics allow one to group or otherwise organize information from the study experience (e.g., a person trying to remember one word from a list of semantically-related items might first think "I need to search for fruits, because I remember studying a list of fruits."). Once this relational information has been reactivated, item-specific characteristics can be considered (e.g., "I remember seeing the word 'apple' because I had one during breakfast, and 'kiwi' because it is fuzzy."). A subsequent analysis of item-differentiating information may lead to the selection of the correct target item (e.g., "I remember generating the memory cue 'fuzzy' and typing it into the computer, so I must have studied 'kiwi."). The reinstatement of both relational and item-specific information at test aids retrieval; effects of one component in isolation are less beneficial (Craik & Jacoby, 1979; Medin, Goldstone, Gentner, 1993). For example, remembering that one generated the memory cue "fuzzy" during study might lead that individual to report either the word "kiwi" or the word "slipper" as a study target if it was the case that no other contextual information was remembered to help determine what "fuzzy" was related to.

Studied items may be accessed in memory during a testing situation as a result of either recollective or reconstructive processes (see Brainerd & Reyna, 2012). In the case of the former, an individual has access to very strong, vivid memory traces for studied items (as might be consistent with distinctive encoding processes) and can therefore retrieve them with relative ease (e.g., "I know I studied the word 'kiwi' and I was supposed to remember it during testing.") In the case of the latter, an individual may recall general characteristics about a relevant study target but not the target itself. In this case, a more thorough search of memory may be required to determine which candidate response in memory is the most likely to be correct (e.g., "I remember seeing 'kiwi,' but what other fruits were on that list? Was 'kiwi' the word I was supposed to remember or was it another list member?"). This type of reconstructive process may be more necessary when both item-specific and relational information are not used together.

One important general criticism of distinctiveness research was summarized by Schmidt (1991): the argument that distinctive items are remembered better because they are distinctive is circular. This circularity can and should be avoided through the use of a strict operational definition of distinctiveness. Researchers have found such a definition difficult to implement given the diversity of experimental designs that have been used and the “dismaying variety” of theoretical conclusions that have been drawn from this research (Schmidt, 1991, p. 539). In lieu of a broad (and non-specific) construct, Schmidt proposed an incongruity hypothesis, which defines distinct stimuli as those that do not fit into a currently-activated cognitive structure. According to this framework, incongruent stimuli are given more attention at encoding than are more typical stimuli, resulting in greater memorability. This framework is consistent with the definition of distinctive processing used by Smith (2006; 2011), and adopted here, which is the processing of difference (incongruent or unique item characteristics) in the context of similarity (relational or organizational item characteristics).

The role of distinctive information on memory performance varies depending upon the degree to which stored item characteristics align with the demands of the memory test and the degree of pre-existing domain-relevant knowledge that enhances both item-specific and organizational processing (Hunt, & Rawson, 2011). Work by Waddill and McDaniel (1998) showed that memory representations can contain many potentially relevant retrieval cues. Cue relevance is dictated solely by the retrieval context; the actual encoding method does not matter, unless the encoded features are useful in a particular retrieval context. Likewise, unique features are useful only when a subset of items have those features. For example, in paired-associates recognition testing, one's memory of an imagined apple cart might lead to the correct recognition of the pairing "APPLE - CART" and the rejection of both "APPLE - TREE" and "GOLF - CART," because no image was generated at study for those pairings. Given that unique items are likely to be processed differently (or simply as being different) from the other items in a set, resultant memory traces may possess unique features that lead to a greater likelihood of retrieval. In this way, encoding and retrieval processes may work together to allow distinctiveness effects to emerge. Waddill and McDaniel (1998) offered an example of distinctiveness in the comparison of two sentences: "The boy found a huge diamond in the jewelry store" and "the boy found a huge diamond in the applesauce" (p.112). The second sentence contains a description of an abnormal circumstance and is therefore encoded and stored in memory as being abnormal as compared to other more normal sentences. At test this unique characteristic would make the second sentence stand apart from other sentences, thereby increasing the likelihood of it being retrieved.

The memorial benefits afforded by distinctive processing result from the differences that exist among items or concepts, in addition to the interaction between relational and differentiating (i.e., unique) information (Hunt & McDaniel, 1993). For example, Epstein, Philips, and Johnson (1975) found that judging differences among similar items and judging similarities among different items increased subsequent memory performance beyond simply making, for example, similarity judgments for similar items. Additionally, a cue-generation effect was investigated by Mäntylä (1986), who required participants to think of a known characteristic for each of 600 unrelated words. The criterion task required the recollection of each target item in response to a self-generated cue from study, which led to cued recall performance of approximately

90% correct. Related work performed by Mäntylä and Nilsson (1988) showed that uniqueness of each self-generated cue was the major contributor of this effect, given the small degree of conceptual overlap between the cues. An experimental condition implemented in Smith and Hunt's (1996) work (discussed in greater detail below) elaborated on these findings. Their administration of novel relational cues (i.e., taxonomic category labels) at test helped participants to reinstate the prior study context and use the unique, item-specific retrieval cues they had generated during the study of taxonomic lists to improve item recall.

According to Schmidt (1991), the benefits of distinctive processing require a perception of item uniqueness that is created in the context of a common base of comparison for all studied items (also see Hunt, 2002). Thus, the type of distinctiveness that one can assess in a particular paradigm depends upon the type of comparison that is made during encoding. Primary distinctiveness results from the comparison of studied items to those that exist in the immediate context (i.e., in short-term working memory), as illustrated by the von Restorff effect. Secondary distinctiveness, on the other hand, requires a comparison between studied items and a more established set of knowledge structures in long-term memory. For example, McDaniel et al. (1995) required participants to generate images to represent bizarre sentences (e.g., “The maid licked ammonia off the table.”) and more normal sentences (e.g., “The maid spilled ammonia on the table.”). When both types of sentences were studied, memory for the bizarre sentences was increased relative to that for more normal sentences. This effect requires participants’ pre-existing world knowledge as the basis for creating distinctiveness (otherwise, bizarre propositions would not stand out). McDaniel and Geraci (2006) made a case for primary distinctiveness being driven primarily by retrieval processes, given that, for example, an isolation effect can be found even when the isolate is presented prior to the presentation of any surrounding context (i.e., other words that would allow an individual to perceive the uniqueness of the isolate). On the other hand, experimental manipulations that limit participants' ability to recognize and encode stimuli as being distinctive (e.g., by presenting lists consisting only of bizarre propositions; see McDaniel, Dornburg, & Guynn, 2005) often reduce or eliminate effects related to secondary distinctiveness.

1.2 Age Differences in Distinctiveness Effects

Although distinctive processing seems to improve memory quality, it may not be the case that both younger and older adults engage in this kind of processing (or benefit from doing so) to the same degree. For example, Basden, Basden, and Bartlett (1993) required younger (aged 16-30) and older (aged 67-91) adults to sort 100 unrelated words into categories and provide labels for those categories prior to a free recall test. Older adults retrieved fewer categories, words in each category, and words overall at test than did younger adults. The administration of cues aided performance but did not decrease the age difference in performance. The authors concluded that older adults may not use relational information as effectively as younger adults to reinstate the study context during retrieval. This is consistent with Hultsch (1974), who showed that as chronological age (measured by decade from the 20s to the 80s) increased, the degree to which individuals learned about the benefits of organizing information following task experience decreased (i.e., increases in such processing were smaller for older adults).

Geraci et al. (2009) also found support for the notion that the encoding of relational information impacts memory performance. In a task wherein participants studied lists of eight semantically-related words, with one of those being unrelated to the others, participants who reported perceiving the relatedness of list items (along with the one unrelated word) were those who exhibited the strongest distinctiveness effect during retrieval. Likewise, Luszcz, Roberts, and Mattiske (1990) concluded that both younger and older adults benefitted from the combination of distinctive and relational information at test in a categorized list task that required participants to either sort items into categories, rate items on pleasantness, or both.

Inconsistent with the above findings, Jackson and Schneider (1982) found that although younger adults (aged 17-19) engaged in more active rehearsal of lists of unrelated words than did older adults (aged 64-83), there was no age difference in the degree of subjective organization used during study and test (e.g., study stimuli were grouped by the first letter in each word, but participants did not cluster items at recall based on this information). Smith (2006, 2011) expressed the idea that older adults may require more support for relational encoding at study and test for them to capitalize on such information during retrieval because they often do not do so spontaneously (e.g., Luo & Craik, 2009). An example of this kind of support would be to require participants to provide taxonomic labels for studied items. Additionally, an experimenter could simply supply those labels to provide additional encoding or retrieval context.

Witte, Freund, and Sebby (1990) offered another clarification regarding these varied findings. Their experiment involved younger (aged 17-42) and older (55-76) adults studying a single list of 30 unrelated words during 6 successive study-test blocks. A free recall test was given following each study sub-block (wherein each item was studied for 3s), and younger adults consistently outperformed their older counterparts. The authors found that younger adults utilized relational encoding to a greater degree than did older adults and that its use increased across study-test blocks more so for younger than older adults. This may be indicative of a greater understanding of task parameters or possibly a greater appreciation of the benefits of organization. The authors indicated that the variety of results obtained in distinctiveness research may be an artifact of list length rather than the type of recall assessed or the operational definition of subjective organization. Their work along with others such as that of Hultsch (1974) used 30 words, whereas research by Jackson and Schneider (1982) used 18 words and Laurence

(1966, as cited by Witte, Freund, & Sebby, 1990) used 16 pictorial stimuli. Shorter stimuli lists may not be sufficient to necessitate a high degree of relational encoding in younger adults, thereby failing to provide an adequate test of age-related differences in relational processing. In other words, younger adults may not need the contextual support that relational encoding would normally provide when studying a brief list, whereas older adults may benefit from its presence.

Older adults may also fail to engage in adequate item-specific encoding. Rabinowitz, Craik, and Ackerman (1982) required younger and older participants to study paired associates and then recall a target word from each pair at test while being cued either with a weak (i.e., non highly-related) cue that had been present during study or a strong (i.e., highly-related) cue that had not been studied along with the target item (also see Koustaal, 2006). Younger adult performance was greater when strong cues were administered, but older adult performance was similar across both conditions. Older adults apparently processed the paired associates at a more general level than did younger adults, who spontaneously made use of more elaborate, distinctive encoding during study despite the relative weakness of the presented cues. Hay and Jacoby (1999) found that this processing deficit could be overcome if older adults were given more study and retrieval time and also given instructions that emphasized the processing of item-specific information. Consistent with this, Rankin and Collins (1986) either provided participants with elaborations for sentence stimuli during study or required participants to generate them. Younger adults who were asked to generate precise (i.e., specific and relevant) elaborations exhibited higher recall of a target item in each sentence than those who were provided with an elaboration, whereas the opposite was found for older adults. Older adults may have failed to process item-specific information during elaboration (perhaps as a consequence of underestimating its importance for memory performance). They may also have failed to integrate item information with the relational (i.e., contextual) information also present at study (perhaps as a result of reduced working memory capacity, e.g., Smith & Engle, 2011), which would make it more difficult to reinstate the study context and then retrieve the correct information at test.

Taking all of the above information into consideration, it seems that older adults generally fail to appreciate or make use of relational information at study to the same degree characteristic of younger adults, even when the stimuli are very supportive of this type of encoding (as when studying semantically-related lists). Older adults may also fail to generate distinctive cues that are as useful as those generated by younger adults in the sense that they are not unique enough to point toward the proper target. More likely, they do not integrate relational and distinctive information during study, which prevents the proper reinstatement of the study context during retrieval. The encoding of context at study may be of key importance especially to older adults, who may rely upon a high degree of both distinctive and relational information at encoding to compensate for age-related declines in recollective ability that would otherwise hinder memory performance (e.g., Smith, 2011).

1.3 The Metacognitive Perspective on Strategy Use

1.3.1 The Role of Performance Monitoring

Metacognition involves monitoring and controlling one's cognitive processes and encapsulates both metacognitive knowledge and experiences (Flavell, 1979). Metacognitive knowledge includes factors such as one's perceptions about the self as a learner (e.g., perceptions of ability), the parameters of the current task, and strategies that may be used to complete that task. Metacognitive experiences provide feedback from current cognitive operations (e.g., pertaining to the encoding or retrieval of information) and whether or not selected strategies are effective. Nelson and Narens (1990) outlined a framework through which the implementation of cognitive processes can change as a result of metacognitive performance monitoring. When applied to multi-trial learning task, for example, a learner who perceives a failure to meeting a current performance goal may change encoding or retrieval strategies to boost subsequent performance. Due to this feedback loop, further performance monitoring may result in another strategic shift or the maintenance of the current strategy, if it is perceived as being effective. One open question is the degree to which metacognitive knowledge relates to the use of distinctive and relational encoding. Is it the case that individuals are able to appreciate the relative benefit of distinctive processing over that of purely relational processing on the quality of a resultant memory? If so, is this knowledge explicit and is it utilized to improve task performance?

1.3.2 Updating Strategy-Specific Knowledge

"Strategy knowledge is declarative knowledge about possible strategies for learning and remembering, including their features and circumstances under which they are likely to be more or less effective in promoting learning" (Hertzog & Dunlosky, 2004, p. 221). Research pertaining to strategy knowledge and learning about strategic effectiveness (i.e., “knowledge updating”) emphasizes the role of monitoring memory outcomes and attributing those outcomes to a specific type of encoding process or strategy (e.g., Dunlosky & Hertzog, 2000). The key components of knowledge updating were outlined in a framework proposed by Dunlosky and Hertzog. This framework involves four assumptions that must be met for knowledge updating to be manifested after task experience. First, different strategies must yield differential memory outcomes. If strategies are not differentially effective, experience in their use would not necessarily lead an individual to adopt one over the other, assuming similar levels of effort are involved in their implementation. When assessing age differences in knowledge updating, it is also important for all age groups to demonstrate a performance difference between strategies, which could be difficult to obtain if, for example, memory performance is universally poor for older adults but not for younger adults. Second, individuals must be capable of accurate metacognitive monitoring. If strategies are differentially effective, this information must be perceived correctly by the learner to capitalize on it. Third, learners must attribute differential performance to the use of different strategies. If a link is not made between encoding method and memory outcome, a learner will be unable to correctly determine why performance varied between items and may simply assume that encoding efforts were generally poor. Finally, learners need to make use of their knowledge about differential strategic effectiveness. Assuming that learners are able to perceive differential performance and attribute it correctly to the use of different strategies, they still may still choose not to use this information (e.g., if a more effective strategy is difficult to implement).

There are a variety of measures that have been used to tap into the products of metacognitive monitoring to assess the relationship between cognitive performance, monitoring, and control processes (e.g., Hertzog & Dunlosky, 2004; Hertzog, Price, & Dunlosky, 2008). Some of these measures are used to assess pre-experimental knowledge (e.g., that which is derived from prior strategic implementation outside of the current task context). These measures are often taken at a global level for each specific strategy (e.g., inquiring about the general effectiveness of a strategy or how many items overall one would expect to remember out of a to-be-studied list). Other measures assess knowledge gains during the completion of an experimental task (e.g., following encoding or retrieval). These measures may be of a global nature (i.e., strategy-differentiated predictions and postdictions), but they can also take the form of a series of item-level assessments such as the judgment of learning (JOL) that asks participants about the likelihood of remembering an item following study but preceding a test experience.

Both global assessments of pre-experimental knowledge and memory performance estimates made between study and test experiences are often poor indicators of actual memory performance (e.g., Lovelace, 1990; Koriat, 1997; Dunlosky & Hertzog, 2000; Hertzog & Dunlosky, 2004). Item-level JOLs, for example, often fail to reflect differences in memory quality that result from the use of different study strategies (Bieman-Copland & Charness, 1994; Hertzog et al., 2008). The component of task experience that appears to increase the accuracy of metamemory judgments is a person's perceptions of memory performance at test. This relationship is evident in the memory for past test (MPT) heuristic (e.g., Finn & Metcalfe, 2007; 2008). This heuristic is relevant when individuals are given multiple opportunities to study information (e.g., in multiple study-test phases for the same items); participants in many cases have been found to rely primarily upon their memories of prior test experiences (i.e., successes and failures in remembering studied items) when deciding how to allocate subsequent study efforts. Individuals are generally able to perceive and remember their previous retrieval performance accurately (e.g., Gardiner & Klee, 1976; Higham, 2002; Finn & Metcalfe, 2008) and adjust their behaviors accordingly. For example, Finn and Metcalfe (2007) found that JOLs made during an initial study phase were less correlated with phase 1 test performance than were JOLs made for the same items during a second study phase. This reflects the fact that people often gain a lot of information about their actual level of learning from a test experience, and that they apply this information when estimating how well they will perform later (for further evidence, see Hertzog, Dixon, & Hultsch, 1990; Lovelace, 1984; Thiede, 1999).

Measures other than JOLs also offer insight into knowledge updating. Item-level confidence judgments (CJs) and globally differentiated postdictions that follow a test are two typical measures utilized in metacognitive research that reflect the learning that takes place during testing. Information related to memory performance can be very valuable for knowledge updating, providing that outcomes can be appropriately linked to specific encoding strategies. One way in which this linkage can be assessed is through the use of global differentiated postdictions, which are thought to require someone to reflect on the overall level of performance associated with each strategy of interest and may reflect the consolidation of item-level CJs (see Hertzog et al., 2008).

Item-level accuracy and CJs are typically highly correlated (e.g., Dunlosky & Hertzog, 2000), but the process of converting accurate item-level performance monitoring into a more global-level inference about strategy effectiveness is not infallible (see Dunlosky & Hertzog, 2000; Hertzog et al., 2008; Hertzog et al., 2009). People may not remember which encoding strategy was used for each item due to interference or memory decay. They will likely not attempt to count instances of item memory successes and failures, either. Instead, they may rely on more abstract heuristics to estimate performance (e.g., sampling from the items immediately available in memory rather than performing an exhaustive search or, alternatively, focusing only on a small subset of possible influences on memory performance). For example, Tullis and Benjamin (2012) found that both younger and older adults began to favor more distinctive low-frequency words over high-frequency words when making item-level memorability judgments across learning trials (reflective of appropriate knowledge updating). This effect was enhanced modestly for participants who made global memory performance postdictions, indicating that these judgments may help people focus on the most stable and relevant cues for retreival, thereby aiding the translation of item-level to global-level performance judgments. Unfortunately, salient cues such as overt stimulus characteristics do not always relate to actual memorability, therefore relying on faulty assumptions about the usefulness of certain cues and how they relate to memory performance or strategic effectiveness can lead to inaccurate judgments (e.g., Koriat, 1997).

1.3.3 Age Differences in Spontaneous Strategy Use

Past research has repeatedly shown that older adults do not learn as quickly or easily as do younger adults in a variety of situations (e.g., Craik, 2006; Luo & Craik, 2008; 2009; McDaniel, Einstein, & Jacoby, 2008; Salthouse, 1996). Age-related deficits in memory performance have been hypothesized to result in part from older adults' less frequent implementation of encoding strategies that have been shown to typically afford higher memory performance (Dunlosky & Hertzog, 1998). For example, normatively effective strategies in associative learning tasks include imagery and sentence generation, whereas less effective strategies include rote repetition or even no rehearsal under conditions when people have high memory confidence and do not feel it is necessary to expend much effort to encode information. The types of tasks that have been used to examine age differences in strategy use often include paired associates learning (e.g., Hertzog, Price, & Dunlosky, 2012). Working memory tasks such as reading or operation span and simple list learning have also been employed (e.g., Bailey, Dunlosky, & Hertzog, 2009) as well as those requiring text comprehension (e.g., Stine-Morrow, Gagne, Morrow, & DeWall, 2004).

Studies comparing strategy use across different age groups reveal an age-related deficit in the spontaneous engagement of effective encoding strategies, which is often referred to as a production deficiency (Kausler, 1994). For example, Murphy, Schmitt, Caruso, and Sanders (1987) found that explicit instructions to perform self-testing during study were required for older adults to actually adopt that effective study strategy during a serial recall task. Also, Dunlosky and Connor (1997) found that older adults failed to optimize their study time allocation in a multi-trial learning experiment, whereas younger adults' study allocation was more highly-related to their previous test performance. Hertzog and Dunlosky (2004) presented data collected prior to the completion of a paired associates task that demonstrated that participants were consciously aware that some strategies are typically more effective than others prior to the task, but older adults did not rate the effectiveness of normatively useful strategies (e.g., imagery) as highly as did younger adults. Such age differences in initial strategy knowledge could impact spontaneous strategy use in a relevant cognitive task.

However, age-related differences in strategy use do not necessarily account for a large proportion of the age-related variance in associative memory performance. Dunlosky and Hertzog (2001) either did or did not inform participants about potential encoding strategies prior to the completion of an associative memory task and found that both younger and older adults used effective strategies following familiarization, whereas only younger adults did so without it. Simply telling older adults that strategies exist can encourage their use above baseline, but such information did not in this case eliminate age differences in memory performance as a result of equalizing strategy use. An age-related deficiency in strategy production is evident in these data, but a utilization deficiency also appears to play a role in memory performance whereby older adults failed to use strategies as effectively as their younger counterparts (also see Hertzog & Dunlosky, 2004). A retrieval deficit may also come into play whereby the products of encoding (e.g., contextual memory cues) may be more likely to be forgotten by older adults, even in the case of similar strategy use. These hypotheses are not central to the current work, but they will be revisited when interpreting the results.

1.3.4 Age Differences in Knowledge Updating

Irrespective of the support that may be required to encourage the initial use of relevant encoding strategies, research has often (but not decisively) shown that both younger and older adults are similarly capable of perceiving the differential effectiveness of such strategies following their use and apply this knowledge to benefit their subsequent task performance (e.g., Dunlosky & Hertzog, 2000; Hertzog & Hultsch, 2000; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Price, Hertzog, & Dunlosky, 2008). As an example, Connor, Dunlosky, and Hertzog (1997) showed that both younger and older adults produced JOLs with equivalent resolution to future test performance whether they were collected immediately after study or following a brief delay. Additionally, item-level and global post-test measures of perceived performance were more highly-correlated to prior recall accuracy than were performance predictions collected prior to testing.

The finding that age spares knowledge updating is not universally supported. Brigham and Pressley (1988) gave participants experience using two differentially effective strategies for learning the meaning of new words. Whereas younger adults showed a strong preference for the more effective strategy following practice, older adults showed no preference for either strategy (which was initially the case for both age groups). One shortcoming of this particular study was that the two strategies were associated with less differentiated performance for older adults, which would have made it more difficult for them to determine which strategy was more effective, but other researchers have also found an age deficit in knowledge updating. For example, Bieman-Copland and Charness (1994) showed that only younger adults were capable of strategy-specific updating in their performance predictions following a test experience in an associative learning task, whereas older adults adjusted their performance expectations globally without considering the strategy that led to a particular memory outcome.

Why is it the case that some studies have found evidence for an age-related decline in knowledge updating, though others have not? Dunlosky and Hertzog (2000) discussed the possibility that different measures of knowledge updating could lead to different conclusions. In a paired-associates learning task the absolute accuracy of estimates (i.e., metacognitive judgments – recall performance) was found to be a poor indicator of updating, unlike relative measures of accuracy (e.g., Pearson correlations between judgments and recall). Matvey et al. (2002) found similar patterns of knowledge updating, and they suggested that older adults may enact deficient inferential processes when translating item-level relationships between encoding strategy and outcome into global estimates of overall strategic effectiveness. Older adults therefore may be able to determine which strategy is best, but the translation of this information into an absolute judgment is imperfect (also see Hertzog et al., 2008).

There are several possible underlying causes of an age-related inferential deficit (Matvey et al., 2002). First, older adults may simply forget which items were recalled correctly when making subsequent global effectiveness judgments. Second, they may not correctly estimate the frequency with which certain strategies led to a positive memory outcome (which could result from a failure to implement a memory search for all tested items and their outcomes). Finally, older adults may anchor their post-test performance based on a general perception of performance, which could, for example, result from experiencing worse memory performance than they were initially expecting. Older adults in this case might adjust performance predictions downward for all strategies in a global fashion (though every strategy may not be given the same rating).

Price et al. (2008) and Hertzog et al. (2009) reduced the age difference in the absolute accuracy of global estimates by exposing participants to testing that was blocked into sets of items that were encoded using the same strategy. This was done to try to reduce the difficulty of making inferences about strategic effectiveness, and blocked testing did largely correct older adults' underestimation of the overall effectiveness of imagery encoding (Hertzog et al., 2009). Mixed testing evidently makes it harder to access and compile information regarding the original encoding strategy used for each item, either at recall or when postdictions are made. The separation of test trials into blocks by study strategy can facilitate the creation of overall strategic effectiveness estimates for items within each block (Hertzog et al., 2008). This blocking effect was not found to be as strong for older adults (Price et al., 2008). One possible reason for this difference is that, as mentioned above, older adults may be less likely to engage in distinctive encoding about study strategies during the process of studying to-be-remembered information and are therefore less likely to access this information during either item-level testing or the construction of strategy-specific postdictions.

Another task characteristic that may be of benefit to knowledge updating is experimenter-guided strategic implementation (as compared to spontaneous strategic selection on the part of each participant). Experimenter-guided strategy use might afford individuals an opportunity to not only gain practice with potentially unfamiliar or rarely-used study strategies, but to also construct an informed opinion regarding their viability in terms of both ease of use and the quality of resultant memory performance in the current task. Hertzog, Price, and Dunlosky (2012) examined the potential for differential knowledge updating in the context of experimenter-guided and participant-chosen strategy use in an initial task phase and resultant strategic implementation during a second study opportunity. They found that supervised experience had a modest effect on the degree of knowledge updating that occurred in their paired-associates learning task, but larger behavioral shifts toward the normatively more effective interactive imagery did follow guided study, with a slightly greater shift occurring for younger than older adults. The authors stated that this may be due to an inertial tendency on the part of older adults to continue using a particular (usually easy-to-implement) strategy in a habitual fashion despite knowing that (1) another viable strategy exists and (2) the alternative strategy typically leads to better memory outcomes (see Hertzog, 2008, for a more elaborate explanation of this behavior). It remains to be seen in the current context (i.e., comparing a distinctive encoding strategy to one focusing solely on relational information) what effect, if any, experimenter-guided strategy experience will have on strategy knowledge and implementation. One might assume that it would lead to enhanced retrieval, at least on the part of older adults, who often benefit from additional encoding support (e.g., Hay & Jacoby, 1999; Smith, 2006; 2011). On the other hand, under conditions of self-selected strategy use, older adults might be able to compensate for potentially lower-quality encoding by simply studying longer to reach a desired level of memorization (e.g., Hines, Touron & Hertzog, 2009).

1.4 Age Differences in Learning About Distinctive Encoding

In line with Dunlosky and Hertzog (2000), both younger and older adults benefit from the use of distinctiveness processing and are generally able to monitor their memory performance, not only at the item-level, but also in a broader sense (especially if the task supports the translation of item-level performance to overall strategic assessments). Under these conditions, it is possible that strategy-specific knowledge updating could occur, but an important question left unaddressed thus far is whether or not individuals become aware of the memory benefits associated with the use of a distinctive encoding strategy. In most cases, distinctiveness experiments have required that participants be assigned to one particular encoding condition or another in a between-subjects fashion, so participants were unable to (1) reveal any expectations regarding the outcomes of different encoding methods and (2) reveal within-person levels of performance for each type of encoding. Metacognitive studies often suggest that effects are more likely to be observed in within-subjects manipulations that allow participants to explicitly contrast benefits of conditions they have seen vary in an experiment (e.g., Carroll & Nelson, 1993). It is therefore of interest in the current investigation to discover how much of a benefit is offered by distinctive encoding, whether this benefit differs between younger and older adults, and how much individuals vary with respect to the benefit received (i.e., how consistent is the effect?). Also, is the difference in memory performance between items studied using distinctive encoding and purely relational encoding strategies suitable to allow participants to (1) appreciate the differential effectiveness of each strategy and (2) encourage a strategic shift toward the use of distinctive processing?

1.5 Hypotheses for the Current Experiment

The current research investigated the following hypotheses. First, older and younger adults' memory performance benefit from the use of a distinctive encoding strategy over a purely relational one (e.g., Smith, 2006; 2011). Second, although older adults' memories would benefit from distinctive encoding, an age-related episodic memory deficit results in lower overall free recall for older adults as compared to younger adults. This memory deficit may not necessarily reflect a failure to remember a target word, but, rather a failure to properly use relational and item-specific information in tandem to differentiate designated targets from other studied items (intrusion errors may therefore be more frequent for older adults). Third, self-initiated use of distinctive encoding would be greater for younger than older adults, given the latter group's more frequent use of more general encoding (e.g., Rabinowitz et al., 1982; Rankin & Collins, 1986; Hay & Jacoby, 1999). Fourth, both younger and older adults would monitor trial-level memory performance (CJs) similarly, but older adults' global differentiated performance postdictions would differ from actual performance more so than those of younger adults (e.g., Dunlosky & Hertzog, 2000; Matvey et al., 2002; Price et al., 2008; Hertzog et al., 2009). This difference was expected to be in the direction of older adults underestimating the effectiveness of a distinctive encoding strategy and (or) overestimating the effectiveness of a purely relational encoding strategy when compared to younger adults. Given a free recall memory test, blocked testing was not possible (as it was in, e.g., Hertzog et al., 2008), and the effect of blocked study may not be sufficient for older adults to overcome their inferential processing deficit. Related to this hypothesis, older adults (unlike younger adults) should display less of a shift from purely relational to distinctive encoding in a second study phase regardless of whether or not they experienced experimenter-guided or unguided study in an initial phase. This age difference in strategic shifting was expected as a consequence of poorer differentiation of global strategic effectiveness ratings on the part of older adults.

A common finding in research on memory and aging is that older adults fail to perform at the level of younger adults in a wide variety of contexts. One factor that could partially explain age differences in memory performance is a failure to utilize highly-effective encoding strategies such as those that focus one's attention on the distinctive properties of to-be-remembered information. The underlying cause of this age difference may be that older adults fail to perceive a benefit (or as great a benefit) associated with the use of a more effective strategy. Both traditional measures of absolute and relative accuracy of strategic knowledge were used to assess any possible age differences in the perception of strategic differences as well as any additional strategy knowledge gained following task experience that might increase the accuracy of those perceptions.

CHAPTER 2

METHOD

2.1 Design

The experiment was a 2 (Age: younger adult, older adult) x 2 (Task Phase: 1, 2) x 2 (Phase 1 Study Type: guided, unguided) design, with age and phase 1 study type as between-subjects variables and task phase as a within-subjects variable. Participants in the unguided condition reported using both distinctive and relational encoding (at least one instance of each for all participants), thereby facilitating a comparison of results by self-reported encoding strategy for both guided and unguided conditions, though the baseline proportions of strategy use differed between groups. The validity of strategic self reports as well as the variability in strategic use are discussed in section 3.1.1.

2.2 Participants

The current sample was made up of 60 younger adults between the ages of 18 and 23 years of age and 58 older adults between the ages of 61 and 78. Typical age effects (all $p < .005$) were found such that older adults reported more years of completed education, scored higher on the Shipley vocabulary test (Zachary, 1986), completed fewer items in the WAIS Digit Symbol task (Wechsler, 1981), and recalled more items during the free recall portion of the WAIS Digit Symbol task (no interactions were reliable; see Table 2.1 for participant characteristics and section 2.3.1 for pre-test task descriptions).

The younger sample was drawn from the Human Subjects Pool of the Georgia Tech School of Psychology, and students received class credit for their participation. Older adults were recruited from the community surrounding Georgia Tech and were given an honorarium of $30. Approximately half of each age group was randomly assigned to each of 2 experimental conditions (unequal cell sizes resulted from dropping participants from analyses, as discussed below; 32 younger adults were guided, 28 younger adults were unguided, 30 older adults were guided, and 28 older adults were unguided). All participants were prescreened for basic health issues that could impede participation using self-reported general state of health, visual impairments, or prescription drug use (especially those that could cause drowsiness or otherwise alter one's state of being, e.g., marijuana). Participants were also required to speak English as their primary language to facilitate stimulus comprehension. Eleven participants' data (4 older adults and 7 younger adults) were not analyzed due to extremely low recall performance (<10% of items recalled), indicative of either a lack of effort (supported by session logs for several of the younger adults) or an undiagnosed (or unreported) memory disorder (which was primarily a concern for older adult participants).

Table 2.1

Participant characteristics by age and phase 1 study type (means and standard errors)

Age	Study Type	N	Age	Education	Shipley	WAIS-C	WAIS-R
Younger	Guided	32	20.03 (0.49)	14.70 (0.23)	14.70 (0.23)	72.24 (2.05)	8.30 (0.28)
Younger	Unguided	28	19.77 (0.50)	14.87 (0.24)	14.87 (0.24)	77.65 (2.11)	8.48 (0.28)
Older	Guided	30	69.83 (0.58)	17.61 (0.28)	17.61 (0.28)	45.17 (2.45)	5.83 (0.33)
Older	Unguided	28	71.80 (0.56)	17.52 (0.27)	17.52 (0.27)	45.32 (2.35)	6.00 (0.32)

Note. N = number of participants in each cell. Age = reported age. Education = reported years of education completed. Shipley = number of correct responses. WAIS-C = correct responses on the WAIS digit symbol test. WAIS-R = correctly recalled symbols on the WAIS digit symbol test.

2.3 Materials and Procedures

2.3.1 Pre-Test Measures

Informed consent was established prior to beginning each experimental session. Participants then completed a brief personal data questionnaire in which they reported basic information about their general health and cognitive function (e.g., the existence of diagnosed memory disorders). Following the questionnaire, participants completed two pencil-and-paper tasks: the Shipley Institute of Living Scale – Revised vocabulary test (Zachary, 1986), and the Digit Symbol Substitution component of the Wechsler Adult Intelligence Scale – Revised (Wechsler, 1981). Following the administration of the above measures, participants began a computer task that was presented on a 15-to-19 inch cathode ray tube (CRT) monitor, and stimuli were presented in a capitalized 14-point Arial font to enhance legibility. Participants were seated at a height and distance from the screen that enhanced their self-reported viewing and comfort. Self-paced task instructions were presented prior to each phase of the computer task, and relevant questions regarding the completion of tasks were addressed by the experimenter either during the session or during the debriefing process, as appropriate.

2.3.2 Criterion Task

2.3.2.1 Experimental Design Selection

The study presented here examined age differences in effects related to primary distinctiveness using an experimental design similar to that employed by Smith and Hunt (1996, 2000). They examined the memory outcomes associated with the encoding of semantically-related word lists. Their typical task involved the presentation of multiple study lists wherein all constituents of each list were of the same semantic category (e.g., "fruits"). Participants were instructed to perform one of two types of encoding during their study of each list, with the criterion memory measure typically being the recall of a single target list member. The first type of encoding required participants to study the top item in each vertically-presented list while focusing on the similarities among the target and other list members. This encoding emphasized relational information that would typically be used to organize items into semantic categories. The second type of encoding required participants to focus on the differences between a single target item and other list members. For example, a person instructed to perform relational encoding might produce a retrieval cue such as “the first item in the list, "apple," is a fruit, like the other items,” whereas a person performing distinctive encoding might produce the following cue: “the first item in the list, "apple," is red, whereas the others are different colors.”

Hunt and Seta (1984) argued that participants in this type of task engage in category-based relational encoding as a result of spontaneously identifying shared properties of list constituents. Consistent with this idea, Smith and Hunt (1996) concluded that additional effort made during distinctive encoding (focusing on the unique properties of a target item following spontaneous relational encoding) led to better memory performance than when participants engaged in relational encoding alone. Given a free recall criterion measure for younger adult participants, distinctive encoding (M = 0.60, SD = 0.09) led to almost twice the memory performance that followed relational encoding (M = 0.35, SD = 0.07). A cued recall task using self-generated cues yielded a smaller, but still sizeable, performance difference (Distinctive: M = 0.97, SD = 0.02; Relational: M = 0.59, SD = 0.07). The benefits of distinctive processing are therefore readily apparent in their sample; the interaction of relational and distinctive information was of great benefit to memory performance.

Smith and Hunt's (1996) participants required more time to generate a distinctive cue than one based on relational information, but their memory performance data indicated that this difference in study time was less important for recall than was the qualitative difference in processing that occurred in each condition. When asked to generate either three relational cues or one distinctive cue at study, for example, overall study time was similar, but memory performance was better for the group that generated the single distinctive cue. Relational information alone did not have a large direct effect on item memorability; instead, it contributed to item memory by establishing the episodic (i.e., spatial, temporal, and semantic) context for other, item-specific, information. Even when prompted by experimenters to only engage in distinctive target encoding, participants benefitted from the spontaneous appreciation of shared item features (i.e., semantic relatedness of list items) that helped to reinstate the study context during testing.

Free recall was selected as the criterion memory outcome for the current experiment following pilot testing of free recall and cued recall (using either a semantic category label or a participant's self-generated distinctive or relational memory cue presented during recall). Cued recall performance led to near-ceiling levels of memory performance by younger adults (>90% accuracy). Younger adults would therefore have been greatly limited with respect to any potential performance improvement to be gained as a result of enhanced strategic knowledge and utilization, and age comparisons of the benefit of distinctive over relational encoding would have also been compromised.

The semantically-related list-learning task employed by Smith and Hunt is ideal for the study of metacognitive knowledge about the benefits associated with distinctive encoding because it allows a clear differentiation between purely relational encoding and distinctive (i.e., relational + differentiating) encoding of to-be-remembered stimuli. After generating a single relational memory cue for each list (i.e., a category label), participants can either generate another such cue or an item-specific cue for a highlighted target item. Strategy-specific performance predictions and postdictions can be assessed along with item-level JOLs and CJs to track not only expectations about future memory performance, but also perceptions of actual prior performance. In the case of experimenter-guided strategy use in an initial study-test phase, participants can gain experience using both strategies back-to-back (in a blocked fashion mimicking the test phase of Price et al., 2008 and Hertzog et al., 2009), which may afford a direct comparison of the quality of encoding offered by each strategy. In the case of self-guided strategy use, participants can apply any prior strategy knowledge they possess, as well as any preference they might have for one method over the other (e.g., a benefit of distinctive encoding would not necessarily lead to its use if it is perceived as being difficult to implement or the benefit is not perceived to be substantial).

In a second study-test phase, participants could apply any strategy knowledge they gained in the first study-test phase. The comparison of strategic implementation between individuals in phase 2 who were assigned to either experimenter-guided or unguided study in phase 1 can potentially reveal how much individuals can learn about these strategies independently (at least, within the current task context). Even if self-guided participation resulted in the implementation of only 1 strategy in phase 1, so participants cannot learn more about the relative effectiveness of each strategy, their test performance (which both younger and older adults can monitor accurately) should provide adequate feedback so that successful performance could lead to a continuation of that strategic approach in phase 2. Conversely, participants might change strategies if their memory performance goals are not met in phase 1.

2.3.2.2 Final Design of the Current Criterion Task

The computer task consisted of two phases, which each included a block of list study followed by a block of free recall testing. Dual non-overlapping sets of categorized lists containing 5 nouns each were drawn from a larger set of 64 categorical norms created by Van Overschelde, Rawson, and Dunlosky (2004). Those authors updated and expanded a previous set of norms created by Battig and Montague (1969). Given the decades that have passed since the construction of the original set of norms, it was prudent to use the more current set. The reasoning behind this decision is made clear by

Van Overschelde et al. (2004, pg. 1), “in the mid-1960s, the waltz was a popular dance, and undergrads wore rubbers on their feet.”

One independent variable was manipulated between-subjects: whether individuals' strategy use was experimenter-guided or unguided. Those in the guided study condition were instructed to use either relational or distinctive encoding for each list (after performing initial relational encoding), whereas those in the unguided study condition were allowed to select their method of study for each list (following initial relational encoding).

Prior to the beginning of the list-learning task, all participants were informed of the existence of two primary study strategies. These instructions were similar to those given to participants by Smith and Hunt (1996, 2000), and described the relational and distinctive encoding methods as well as provided examples of each. Although past research such as Dunlosky and Hertzog (1998; 2001) showed that such strategic information can lead to the adoption of encoding strategies by older adults (above a baseline of spontaneous use), pre-study strategy-specific performance predictions would have been impossible to obtain without these instructions. Important information about knowledge updating that occurred during phase 1 would have otherwise been missed. The question of how this design decision may have impacted the results is therefore left for a follow-up study.

To make each strategy more accessible to participants, relational encoding was labeled as "shared" and distinctive encoding was labeled as "unique" in the instructions and within the study portion of the task. The procedure for relational encoding was explained as follows: “When you are studying the target word using the SHARED strategy, you need to focus on the way(s) in which all of the items in the current list are similar to each other. For example, viewing a list with a target word BANANA and non-target words APPLE, STRAWBERRY, GRAPE, and ORANGE might lead someone to think of the word FOOD because all of the words in the list are foods. Once you have thought of one appropriate word, type it in the box and press ENTER to continue." The encoding procedure for distinctive encoding was explained as follows: "When you are studying the target word using the UNIQUE strategy, you need to focus on the way(s) in which the target word in the current list is different from all of the other words in the list. For example, viewing a list with a target word BANANA and non-target words APPLE, STRAWBERRY, GRAPE, and ORANGE might lead someone to think of the word YELLOW because only bananas are that color. Once you have thought of one appropriate word, type it in the box and press ENTER to continue."

After instructions were read, participants received two practice trials for each strategy (that also used semantically-related lists drawn from Van Overschelde et al., 2004) to check their understanding of the task requirements. These trials were formatted exactly like study trials in the rest of the task, but instructions screens before and after the practice trials emphasized that these four trials were not part of the main task. During each trial, all participants were asked to think of a category label that was appropriate for all of the items in the list to ensure that each list member was processed (i.e., they performed relational encoding). Following their self-paced response, one of the words in the list (chosen at random) was increased in size to a 20-point font, and its font was emboldened to make it stand out further from the other words to ensure participants knew which word to focus on during the second part of each trial. Between the first and second screens (after the first word was typed and before the target was emphasized) participants in the guided study condition were either asked to type one word that represented something unique about the target item (the screen flashed "UNIQUE" for 1000ms) or type one word that represented a characteristic shared between that item and the others in the list (the screen flashed "SHARED" for 1000ms). Participants in the unguided condition were asked to type one word that would help them think about the target word (the screen flashed “STUDY” for 1000ms). Following the self-paced generation of this potential memory cue, participants entered a JOL to assess their perceived likelihood of remembering the target item approximately 10 minutes later, on a continuous scale of 0 to 100 percent memory confidence.

To ensure an understanding of both strategies during practice, experimenters watched each participant closely to ensure that their category labels (which could be made up of several words) and memory aid words (the single words typed following instructions to encode the study target in a particular way) were consistent with instructed strategy use. If it appeared that participants did not understand the instructions, their practice was interrupted and the instructions were clarified. Additionally, even after appearing to understand the instructions, participants were asked if they had any additional questions about the two strategies before they were allowed to begin the criterion task. If confusion persisted, participants performed strategy practice again until they understood precisely what was being asked of them. Generally, any confusion was usually resolved during a second practice opportunity (in a few cases, a third was required prior to participants understanding the instructions, which typically coincided with an "ah ha!" moment). Following the completion of practice and subsequent instructions, participants were presented with a fixation cross (for a duration of 500ms), followed by the first study list.

During study, guided lists were grouped by encoding strategy such that participants were exposed to 4 blocks of 8 lists, ordered as shared, unique, shared, unique or presented in a counter-balanced manner. There were no differences in study time, memory performance, or strategy use between the two counterbalanced conditions, so data were collapsed for all analyses. Unguided study lists in Phase 1 were presented in blocks of 8 lists, but were not presented along with any encoding instructions other than "STUDY." To facilitate the adherence to study instructions in the guided condition, each block of lists was preceded by a reiteration of strategy-appropriate study instructions corresponding to presentation block.

Following the completion of the first study phase, all participants completed a 5-minute filler task requiring them to compare a series of letter pairs presented on the computer screen and determine if they were the same letter or two different letters. This 5-minute delay should compensate for common recency effects (Atkinson & Shiffrin, 1968) that could afford disproportionally greater recall of list members that were presented near the end of study, and this delay was consistent with Smith and Hunt's (1996; 2000) previous work.

During Phase 1 test, participants were given 8 minutes to recall as many target words as possible from each list and were instructed to only report the target words that were presented during the previous study phase. After entering each word, participants were required to give a confidence judgment for their response on a scale from 0 to 100 percent memory confidence (i.e., "How confident are you that your previous response was correct?"). These data can be used to assess the degree to which item-level confidence translated into global confidence and effectiveness ratings for each strategy (global estimates of strategic knowledge are discussed in detail below).

A second study-test phase followed the completion of Phase 1 testing. Unlike Phase 1, the presentation of stimuli was not blocked into sets of 8 lists, but consisted of one uninterrupted block of 32 lists. The purpose of the second phase was to assess whether or not participants utilized any knowledge that they acquired in Phase 1 about strategic effectiveness. This utilization may be present in the form of more frequent implementation of one strategy over another, or more differentiated metamemory judgments.

Phase 2 testing began after the completion of another 5 minutes of the above-mentioned filler task, and testing took the same form as in Phase 1. Following its completion, all participants completed a computerized post-task questionnaire meant to assess their strategy implementation throughout the experiment. Participants were presented with each study list with the appropriate target item highlighted in bold to reinstate the study context as completely as possible. The memory aid word that they generated at the time of study was also presented along with the list. Participants were asked whether or not the memory aid word was an example of a SHARED or UNIQUE property of the target, and they entered their response for each list into a text box.

Finally, to assess participants' knowledge, or perceptions, pertaining to each of the two encoding strategies of interest, a brief strategy assessment questionnaire was completed by participants prior to study and following test in both phases (following a similar procedure used by Hertzog & Dunlosky, 2004). This questionnaire asked the following before study: (1) "On a scale of 0 to 100%, how confident would you be that you would remember a word that you studied using the SHARED strategy?"; (2) "On a scale of 0 to 100%, how confident would you be that you would remember a word that you studied using the UNIQUE strategy?" (3) “On a scale of 0 to 100%, how confident would you be that you would remember a word that you studied using any strategy?” (4) “On a scale from 1 (ineffective) to 10 (highly effective), rate the effectiveness of the SHARED strategy.”; (5) “On a scale from 1 (ineffective) to 10 (highly effective), rate the effectiveness of the UNIQUE strategy.”

Following test, participants were asked: (1) "What percentage of items studied using the SHARED strategy did you remember correctly at test?"; (2) "What percentage of items studied using the UNIQUE strategy did you remember correctly at test?"; (3) “What percentage of items studied did you remember correctly at test?” (4) “On a scale from 1 (ineffective) to 10 (highly effective), rate the effectiveness of the SHARED strategy.”; (5) “On a scale from 1 (ineffective) to 10 (highly effective), rate the effectiveness of the UNIQUE strategy.” These global estimates of strategic expectations, success, and effectiveness were used to assess knowledge updating in an overall sense—did participants' experiences using particular study strategies inform them about their usefulness, and did their ratings reflect accurate performance monitoring?

One caveat worth mentioning is that participants may have only used one of the possible study strategies under conditions of unguided study (e.g., if older adults over-emphasized the importance of item-level information and failed to appreciate the importance of relational information), so the comparison of strategic effectiveness between individuals who used only one strategy and those who used both may not be appropriate. For example, those who only used one strategy may have based their ratings for the other strategy on information that was not derived from the current task (e.g., they may have perceived a strategy similar to one described to them in this experiment as ineffective when they used it in the past for a different task, whereas in the current experimental task, they might feel differently after some practice). As a consequence of such issues, participants who exhibited unitary strategic implementation under conditions of unguided study during phase 1 would have been excluded from analyses related to knowledge updating effects. However, all participants reported using both strategies, so this does not appear to be a major issue with the current data set.

CHAPTER 3

RESULTS

SAS PROC MIXED was used to examine differences in mean levels of objective and subjective performance indices. Unless otherwise specified, all analyses were structured as 2 (Age: younger adult, older adult) x 2 (Task Phase: 1, 2) x 2 (Phase 1 Study Type: guided, unguided) x 2 (Participant-Reported Study Strategy: distinctive, relational) design, with age and study type as between-subjects variables and task phase and study strategy as within-subjects variables. Effect sizes were computed as a variation of Cohen's (1988) *d*, such that $d^* = [(M_1 - M_2)$ / SQRT(pooled variance estimate)]. The pooled variance estimate was calculated as the unweighted average of the error variances for within-subjects factors for random effects in PROC MIXED, and *d** can be framed as the number of standard deviations that separate the two means. Cohen (1988) suggested benchmarks of 0.2, 0.5, and 0.8 for small, medium, and large effect sizes, respectfully.

3.1 Study Measures

3.1.1 Strategic Implementation

<u>3.1.1.1 Validation of Strategy Reports</u>

Self-reports of encoding strategies were validated through a procedure in which both the primary investigator and an undergraduate technician independently coded all participant-generated memory aid words as being probable examples of either a distinctive or a relational encoding process. The disagreements (17 of 7562 responses, or approximately 0.2%) were resolved through discussion. The results were encouraging in that they largely indicated a match between self-reported strategy and experimenter-coded strategy. There was an age difference in the proportion of trials for which the self-reported and experimenter-coded strategies corresponded, *F*(1,118) = 8.02, *p* = .005, *d** = 0.29, with more matches for younger adults (*M* = 0.88, *SE* = 0.02) than older adults (*M* = 0.82, *SE* = 0.02). Those in the guided condition (*M* = 0.82, *SE* = 0.02) also displayed a lower proportion of matches than did those in the unguided condition (*M* = 0.88, *SE* = 0.02), *F*(1,118) = 5.29, *p* =.023, *d** = 0.24, indicating that people who were instructed to use a particular strategy might have reported one strategy while actually using another.

Purposeful deception seems unlikely because the strategy reports were made retrospectively after both study-test phases were completed. During strategic reporting participants were not told which strategy they were supposed to have used on a given list (when applicable); only the 5-item list (with the target item highlighted) and participant-generated memory aid word were available as strategic indicators. Instances of unintentional inaccurate strategic reporting were minimized as a consequence of careful strategic orientation prior to the beginning of the task (described above in section 2.3.2.2). Mismatches between participant-reported and experimenter-coded strategies occurred at similar levels in both directions, with approximately 60% of mismatches being due to a self-reported use of distinctive encoding that appeared as relational to experimenters. Due to the high degree of idiosyncrasy that was likely to be implemented when generating memory aid words, self-reported strategies were used for all upcoming analyses.

3.1.1.2 Self-Reported Strategy Use

Contrary to expectations, the overall proportion of use for a distinctive encoding strategy was only marginally larger for younger (*M* = 0.82, *SE* = 0.02) than older adults (*M* = 0.76, *SE* = 0.02), *F*(1,118) = 2.50, *p* = .117, *d** = 0.11. Given prior research results (e.g., those discussed by Smith, 2006), this difference was expected to be statistically reliable. It may have been the case that exposing participants to both types of encoding strategies and enabling their comparison might have encouraged the use of distinctive encoding beyond the expected baseline (this idea will be expanded further in the discussion). The overall proportion of distinctive encoding did increase from phase 1 (M = 0.74, SE = 0.02) to phase 2 (M = 0.84, SE = 0.02), $F(1,118) = 46.43$, $p < .001$, $d^* = 0.29$. A change in the rate of strategy use was expected, given that in phase 1, half of participants (those in the guided condition) were instructed to utilize each strategy equally. A larger change in use occurred for younger adults (Phase 1: M = 0.74, SE = 0.02; Phase 2: M = 0.89, SE = 0.03; d^* = 0.44) than older adults (Phase 1: M = 0.73, SE = 0.02; Phase 2: M = 0.79, SE = 0.03: d^* = 0.18), $F(1,118) = 9.76$, $p = .002$, due to a higher degree of distinctive encoding on the part of younger adults in phase 2. This pattern was entirely driven by shift to greater distinctive strategy use by those in the guided condition (Phase 1: M = 0.61, SE = 0.02; Phase 2: M = 0.85, SE = 0.03; d^* = 0.71) compared to the unguided condition (Phase 1: M = 0.86, SE = 0.03; Phase 2: M = 0.83, SE = 0.03; d^* = 0.09), $F(1,118) = 69.54$, $p < .001$. Again, this was expected due to the nature of guided study in phase 1.

A reliable Age x Condition x Phase interaction was found, $F(1,118) = 10.67$, $p = .001$ (see Figure 3.1). Whereas younger adults used distinctive encoding on approximately 90% of trials when their encoding selection was not constrained by experimenter instructions, older adults did so at a rate of about 80%. This usage pattern in phase 2 was unaffected by guided study in phase 1. It appears that guided study did little more than place a constraint on spontaneous strategic use (a constraint that was adhered to more so by younger than older adults). Once that constraint was removed, encoding strategies were implemented at a similar rate for both previously-guided and unguided participants. Given this information, it appears that participants either possessed beliefs regarding strategic effectiveness prior to entering into this task or these beliefs were constructed as a consequence of reflection induced by strategic instructions. This idea will be expanded upon later.

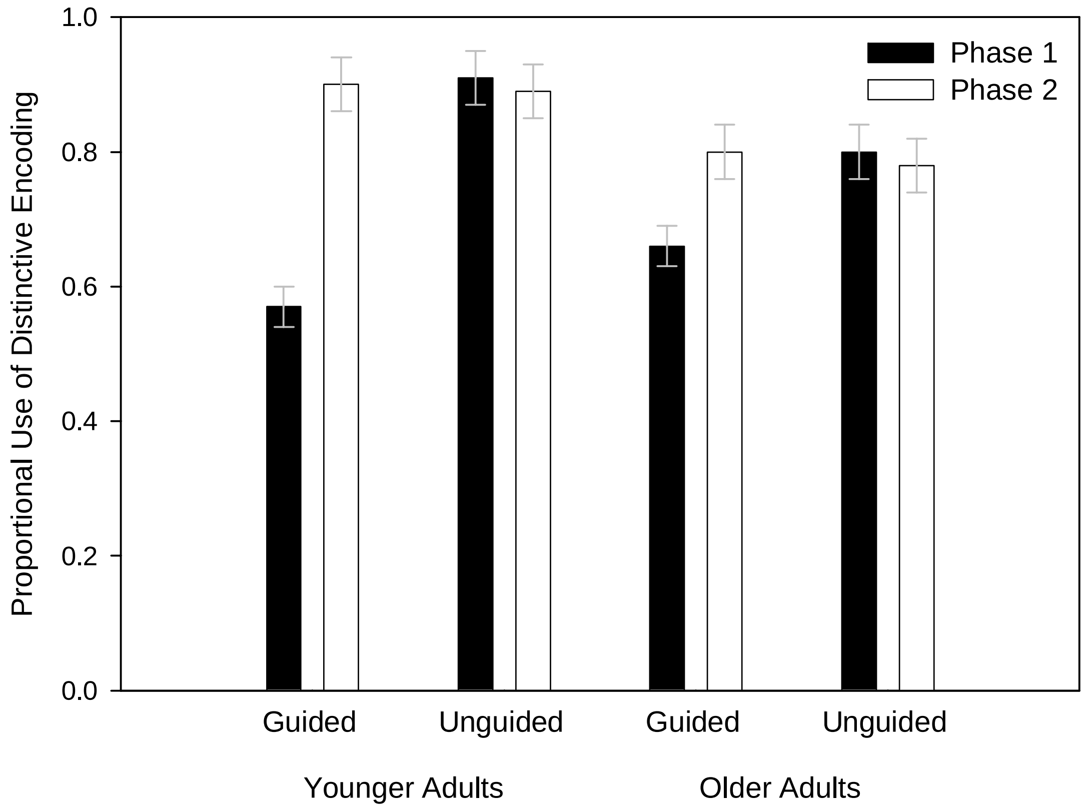

Figure 3.1. Self-reported strategy use by age, condition, and task phase (means and standard errors)

3.1.1.3 Consistency of Strategic Implementation

Pearson product-moment correlations were calculated to assess the stability of individual differences in strategy use between phases 1 and 2. When participants were unguided in both phases, younger adults ($r = .92$, $p < .001$) and older adults ($r = .97$, $p < .001$) used the distinctive encoding strategy very consistently. Guided younger adults exhibited more variability in strategic shift than did their unguided counterparts ($r = .16$, $p = .40$). However, guided older adults' strategy use remained more consistent between phases relative to that of younger adults ($r = .79$, $p < .001$). This stronger correlation resulted from older adults' lower degree of adherence to study instructions in phase 1 relative to that of younger adults, typically favoring the use of distinctive encoding, which aligned more closely with their study behavior in phase 2. Older adults were therefore more likely to display an inertial tendency in strategy use between phases, consistent with paired associates data of Hertzog et al. (2012). The current results provide evidence that older adults may maintain consistent use of a normatively effective strategy (whereas their older adults in Hertzog et al., 2012, persisted in rote encoding--a normatively ineffective strategy). Older adults' strategic perseverance may not always be damaging in terms of task performance if their chosen strategy is relatively effective in a particular task--whether or not they are aware of the strategy's effectiveness.

3.1.2 List-Specific Study

The fluency measures discussed below allowed for an examination of some components of strategic regulation beyond simple strategy choice. One particular relationship that was examined explicitly was that between encoding fluencies and judgments of learning. Encoding fluency has been shown to influence judgments of learning in paired-associates tasks (e.g., Begg et al, 1989; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Koriat & Ma'ayan, 2005). However, encoding fluency is not always indicative of the actual quality of learning unless it relates to qualities of the to-be-learned information (e.g., the concreteness of paired-associates; Begg et al., 1989; Hertzog et al., 2003) that may make encoding harder but might also enhance its quality. In the context of the current experiment, it might be the case that guided study would lead to longer encoding times as a result of constraining participants' natural, or spontaneous, strategy use. However, the greater study effort of guided participants might be rewarded with superior memory performance.

<u>3.1.2.1 Category Generation Time</u>

Category generation RT represents the amount of time a participant utilized when (a) thinking about the ways in which list members were similar while (b) making a determination about the most appropriate commonality to use as the basis of the reported semantic category and, possibly, (c) rehearsing that category. This measure was intended to represent relational encoding time, but it also included the amount of time required to physically type the category label and press ENTER. It is somewhat atypical to analyze processing fluency controlling on other factors that might influence it (e.g, initial processing difficulty, as in Koriat et al., 2006), likely because of the potential for bi-directional relationships between such variables (e.g., difficult words might require more encoding time to reach a desired quality threshold, and a long or effortful encoding experience might likewise relate to a higher encoding difficulty rating). To control for the variability in measured RT introduced by typing speed in the current work, two effects were included in the current ANOVA beyond those specified in the standard equation: the length of the typed category in letters and the age interaction with typed category length (based on the assumption that older adults type more slowly on average than younger adults). Category label generation time, unsurprisingly, increased with label length, $F(25, 3097) = 22.02$, $p < .001$, and older adults (Length of 5: $M = 11.26$s, $SE = 0.73$; Length of 20: $M = 21.94$, $SE = 3.16$; $d^* = 0.55$) exhibited a slightly greater degree of slowing than did younger adults (Length of 5: $M = 5.43$s, $SE = 0.77$; Length of 20: $M = 12.61$s, $SE = 1.30$; $d^* = 0.50$), $F(23, 3097) = 6.07$, $p < .001$.

After controlling for the length of category labels, younger adults ($M = 6.79$s, $SE = 0.52$) generated labels more quickly than did older adults ($M = 13.42$s, $SE = 0.53$), $F(1,446) = 149.77$, $p < .001$, $d^* = 0.31$. Overall category generation speed increased from phase 1 ($M = 10.67$s, $SE = .40$) to phase 2 ($M = 9.53$, $SE = 0.41$), $F(1,3104) = 39.01$, $p < .001$, $d^* = 0.07$, reflecting common practice-related improvements. Category generation time varied between distinctive ($M = 9.75$s, $SE = 0.39$) and relational encoding strategies ($M = 10.45$, $SE = 0.42$), $F(1,3176) = 7.19$, $p = .007$, $d^* = 0.19$, with the difference being driven solely by those in the guided condition (Distinctive: $M =$

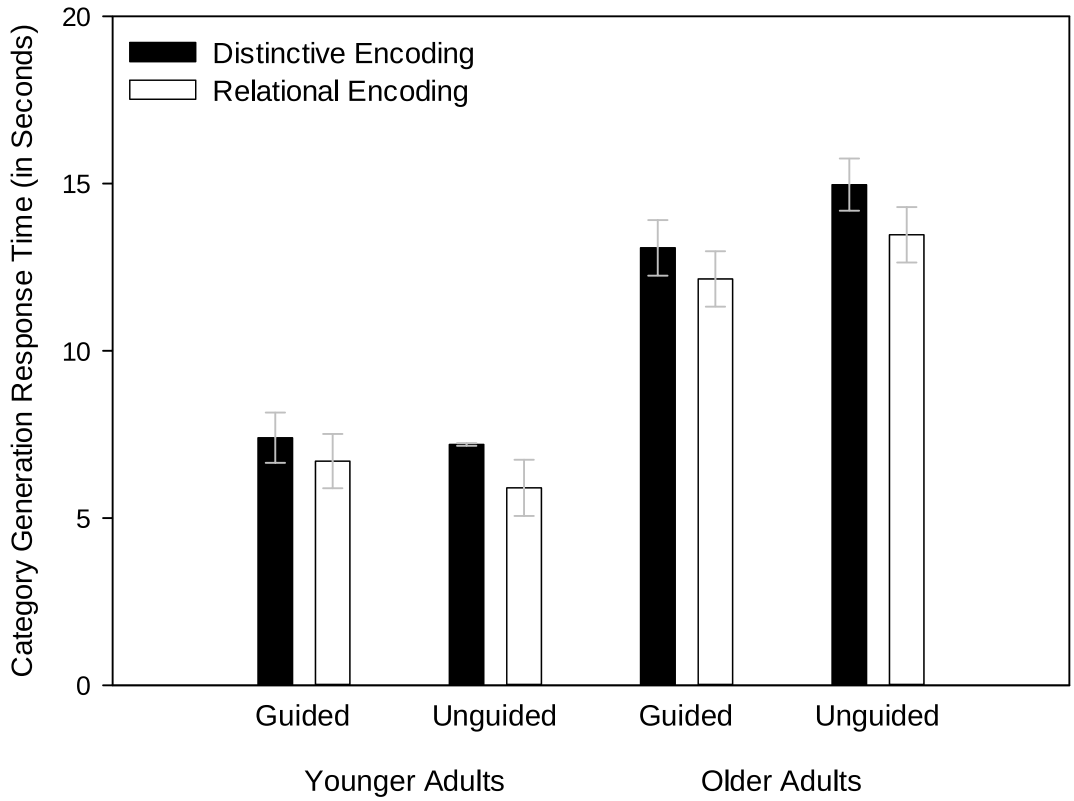

Figure 3.2 Category generation response time in seconds by condition and study strategy (means and standard errors)

9.93s, *SE* = 0.54; Relational: *M* = 11.33s, *SE* = 0.58; *d** = 0.23) rather than the unguided condition (Distinctive: *M* = 9.57s, *SE* = 0.57; Relational: *M* = 9.58s, *SE* = 0.62; *d** = 0.01), $F(1,3177) = 6.79$, $p = .009$. Perhaps when participants were instructed to generate a relational memory aid word after a (relational) category label, they considered more dimensions along which similarities existed prior to settling on a particular label. This is possible because participants were informed of the guided memory aid word generation strategy at the beginning of each block of 8 study lists and had an opportunity to use this information to inform their category label decisions.

<u>3.1.2.2 Memory Aid Word Generation Time</u>

Memory aid word generation RT represents the amount of time a participant utilized when engaging in the focused study of a particular experimenter-chosen target item in relation to the other list members. This measure can be thought of as representing the time spent either (a) thinking of additional ways in which a target was similar to other list members (i.e., engaging in further relational encoding) or (b) thinking of a way in which a target differed from other list members (which, combined with the previous category generation, resulted in distinctive encoding). As with category generation RT, this measure included the amount of time required to physically type the memory aid word and press ENTER. To control for the variability in measured RT introduced by typing speed, the length of a memory aid word and the Age x Length interaction were included in this ANOVA. Memory Aid generation time increased with word length, $F(13, 2388) = 11.27$, $p < .001$, and older adults (Length of 5: $M = 18.52$s, $SE = 1.15$; Length of 10: $M = 23.34$s, $SE = 1.61$; $d^* = 0.17$) again exhibited a greater degree of increase than did younger adults (Length of 5: $M = 11.87$s, $SE = 1.14$; Length of 10: $M = 12.87$s, $SE = 1.44$; $d^* = 0.13$), $F(12, 2388) = 4.20$, $p < .001$.

After controlling for the length of the typed entry, younger adults ($M = 10.81$s, $SE = 1.07$s) were found to have generated Memory Aid words almost twice as fast as did older adults ($M = 19.77$s, $SE = 1.08$), $F(1,260) = 69.82$, $p < .001$, $d^* = 1.09$. Unguided participants ($M = 13.14$s, $SE = 1.05$) were also faster than those in the guided condition ($M = 17.45$s, $SE = 1.05$), $F(1,131) = 11.66$, $p < .001$, $d^* = 0.33$, which was expected given that a particular list target might not be most easily encoded with an experimenter-instructed strategy. Word generation sped up overall from phase 1 ($M = 16.49$s, $SE = 0.87$) to phase 2 ($M = 14.09$s, $SE = 0.91$), $F(1,2395) = 67.39$, $p < .001$, $d^* = 0.17$, likely due to simple practice effects, in addition to removing the encoding restriction from participants who had been guided in phase 1. The Age x Condition x Phase interaction was reliable, $F(1,2394) = 5.46$, $p = .020$ (see Figure 3.3). Guided older adults exhibited a

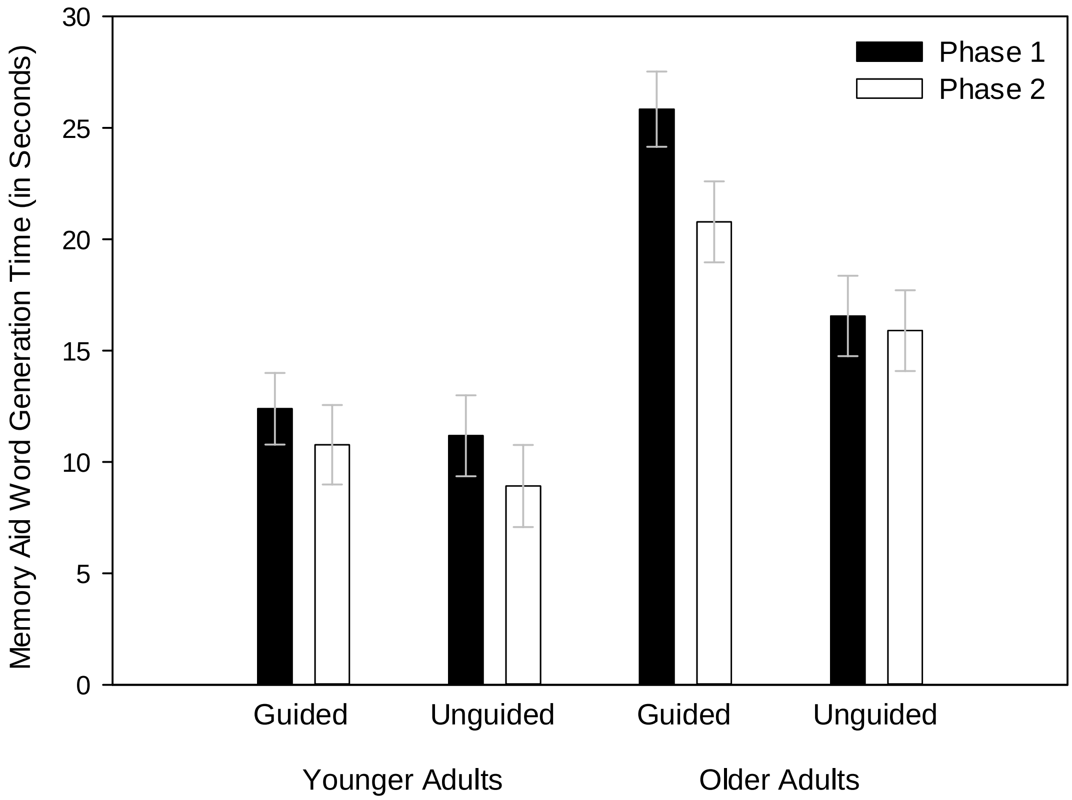

Figure 3.3. Memory aid word generation time in seconds by age, condition, and task phase (means and standard errors)

superior speed increase between phases compared to any other group. Previously-guided older adults did not, however, reach the same generation speed as previously-unguided participants in phase 2, perhaps resulting from a lack of experience selecting their own study strategies. This effect may also be explained by more thoughtful consideration of strategies in phase 2 on the part of previously-guided older adults, who received more practice using both in phase 1 than did their unguided counterparts.

Memory Aid generation speed differed between distinctive (M = 11.82s, SE = 0.84) and relational (M = 18.76s, SE = 0.94) encoding strategies, $F(1,2458) = 82.86$, p <

.001, d^* = 1.04; it was harder to generate a relational word after generating an (also relational) category label. This effect was much larger for those in the guided condition (Distinctive: M = 13.10s, SE = 1.16; Relational: M = 21.79s, SE = 1.28; d^* = 0.78) than the unguided condition (Distinctive: M = 10.55s, SE = 1.22; Relational: M = 15.73s, SE = 1.37; d^* = 0.42), $F(1,2459) = 6.25$, $p = .013$, further supporting the idea that assigning encoding strategies increased encoding difficulty. The Age x Condition x Strategy interaction was reliable, $F(1,2459) = 5.02$, $p = .025$ (see Figure 3.4). The time cost associated with relational encoding was similar between younger adults in both conditions and unguided older adults. On the other hand, guided older adults clearly were at a disadvantage when they were told to generate a relational memory aid word after generating an also-relational category label. This apparent lack of encoding flexibility may have had other consequences as well; if older adults found it more difficult to generate cues than younger adults (an argument indirectly supported by relevant RTs), the quality of their generated mediators might have also suffered. This possibility will be evaluated below using free recall data.

3.1.3 Judgments of Learning

3.1.3.1 Mean Judgments of Learning

Judgments of learning (scaled from 0 to 100%) collected following the generation of a category and memory aid word for each list were collected primarily to assess any differences in initial memory confidence based upon age and encoding strategy. Information related to JOLs collected following phase 2 study were also included in the analyses for the sake of completeness. Judgments did not vary with age, $F(1,118) = 0.09$, $p = .761$. However, confidence decreased from phase 1 (M = 62.41, SE = 1.82) to

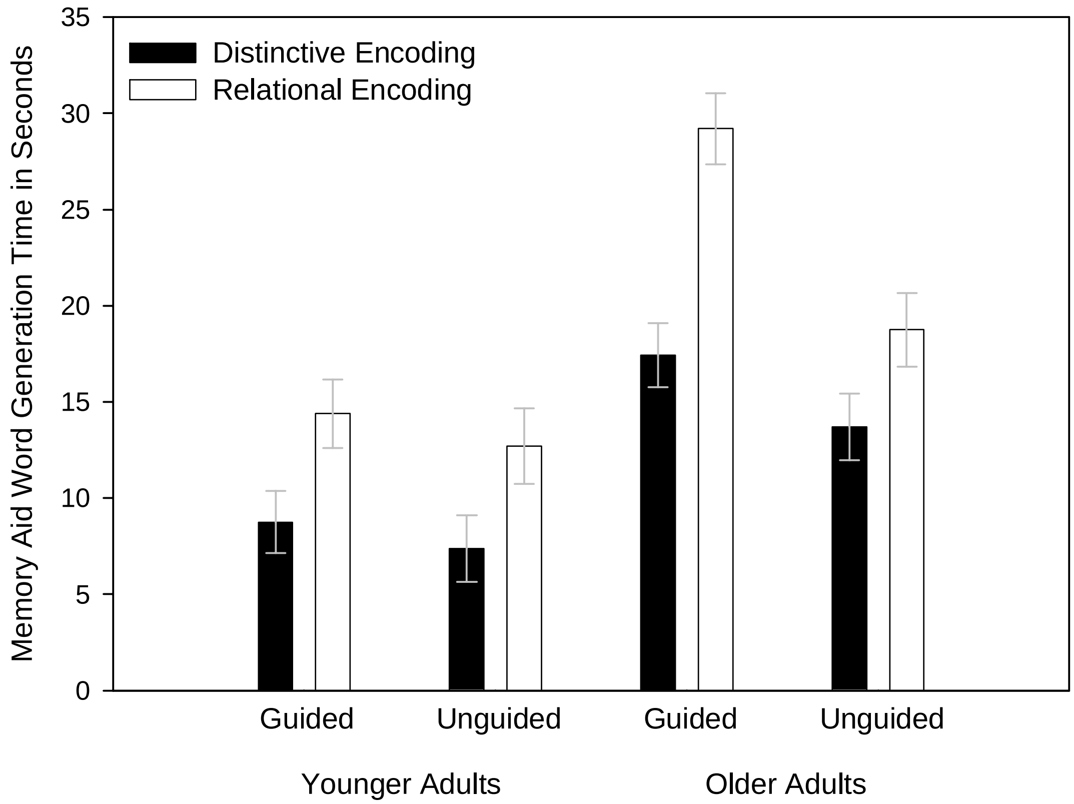

Figure 3.4. Memory aid word generation time by age, condition, and encoding strategy (means in seconds and standard errors)

phase 2 (M = 52.98, SE = 1.86), $F(1,293) = 65.66$, $p < .001$, d^* = 0.82, likely reflecting implicit performance feedback from the first test experience (i.e., participants perceived recalling fewer items than they had initially anticipated). Confidence decreased more so for younger (Phase 1: M = 63.41, SE = 2.56; Phase 2: M = 50.92, SE = 2.61; d^* = 1.10) than older adults (Phase 1: M = 61.40, SE = 2.60; Phase 2: M = 55.05, SE = 2.64; d^* = 0.56), $F(1,293) = 6.96$, $p = .009$, possibly due to a more accurate perception of the difference between expected and actual memory performance in phase 1 (which would be afforded by better young adult episodic memory). Alternatively, this could have resulted from a greater degree of initial overconfidence on the part of younger adults.

Post-study confidence was higher following distinctive (*M* = 62.77, *SE* = 1.80) than relational (*M* = 52.62, *SE* = 1.89) encoding, *F*(1,299) = 71.47, *p* < .001, *d** = 0.89, so participants appeared to appreciated the normative superiority of the distinctive encoding strategy at the item-level. This difference was larger for younger (Distinctive: *M* = 63.54, *SE* = 2.53; Relational: *M* = 50.78, *SE* = 2.66; *d** = 1.12) than older adults (Distinctive: *M* = 61.99, *SE* = 2.57; Relational: *M* = 54.46, *SE* = 2.69; *d** = 0.66), *F*(1,299) = 4.74, *p* = .03, due to inflated confidence in the relational strategy on the part of older adults. The difference in confidence between strategies was also larger for those in the guided condition (Distinctive: *M* = 62.83, *SE* = 2.48; Relational: *M* = 49.27, *SE* = 2.59; *d** = 1.19) than the unguided condition (Distinctive: *M* = 62.71, *SE* = 2.61; Relational: *M* = 55.98, *SE* = 2.75; d* = 0.15), *F*(1,299) = 8.10, *p* = .005, reflecting the possibility that strategic guidance highlighted the superiority of distinctive encoding (or, conversely, the inferiority of relational encoding).

The Age x Condition x Strategy interaction was reliable, *F*(1,299) = 4.57, *p*= .033 (see Figure 3.5). Younger and older adults rated their confidence in the distinctive encoding strategy similarly regardless of condition. However, confidence in the relational strategy was lower for younger adults who experienced guided study compared to those who did not, with the latter group expressing similarly high confidence to that of both older adult groups. These results indicate that participants did not fully appreciate the degree to which relational encoding might hinder memory performance at the item-level. Younger adults, however, were able to do so to a greater extent when task conditions highlighted study strategies and (possibly) created a more accessible association between strategy and memory outcome.

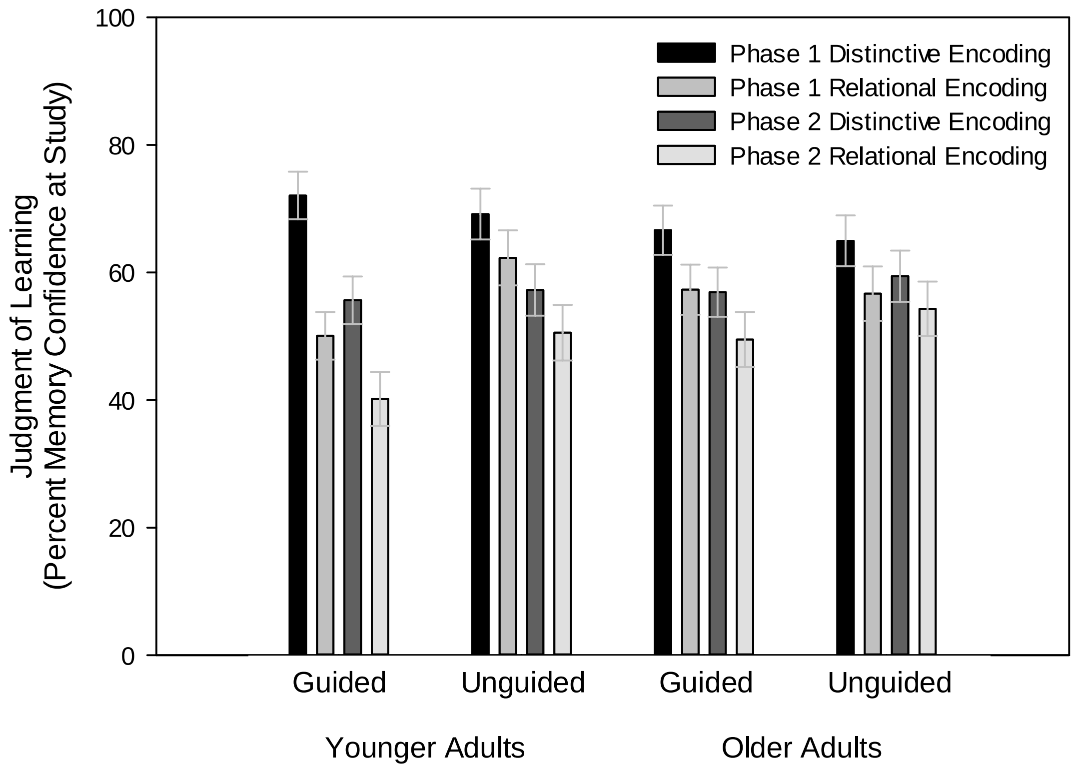

Figure 3.5. Judgment of learning by age, condition, and encoding strategy, and task phase (means and standard errors)

<u>3.1.3.2 Multi-level Regression Predicting Phase 1 Judgments of Learning</u>

Differential study confidence was found between encoding strategies in both phases 1 and 2. These results stand in contrast to those pertaining to paired associates tasks, which commonly reveal an insensitivity of phase 1 JOLs to differential strategic effectiveness (e.g., Bieman-Copland & Charness, 1994; Hertzog et al., 2008; Hertzog et al., 2009). It may have been the case that blocking study lists by encoding strategies highlighted their differential effectiveness. It might also have been the case that encoding fluencies, which were found to differ greatly between strategies, influenced participants' perceptions of the strategies (as discussed above at the beginning of section 3.1.2).

A multi-level regression model was constructed using SAS PROC MIXED (Littell, Milliken, Stroup, & Wolfinger, 2000) to assess which factors predicted phase 1 JOLs. Within-person (Level 1) and between-person (Level 2) effects were estimated in the model (see Singer, 1998). The simultaneous estimation of both kinds of effects allowed the regression model to address questions such as "were higher encoding RTs predictive of a lower JOL in phase 1?" along with "did people who exhibited higher average encoding RTs also have lower than average JOLs in phase 1?" If these within-person and between-person influences are not identified and evaluated in tandem they might contaminate each other (e.g., Snijders & Boskers, 1999).

The initial regression model predicting phase 1 JOLs included the item-level predictor variables of category label length, memory aid length (intended to be simple measures of encoding complexity), category generation RT, and memory aid word generation RT, and encoding strategy. It also included person-level predictor variables for these effects, and the person characteristics of age group and guidance condition. Relevant potential interactions of item-level effects with age and feedback condition were also included in the initial model. Nonsignificant effects were trimmed carefully from the initial model in order to generate a final model, which is presented in Table 3.1

Despite the fact that all regression effects in the final model were evaluated simultaneously in SAS, each effect presented in Table 3.1 will now be discussed in isolation in order to facilitate the an examination of the relative impact of each effect on phase 1 JOLs. First, the intercept indicates a mean JOL of approximately 56% for older adults (dummy variables coded older adults and the guided condition as having 0 weights; estimated effects for other groups are scaled against this reference point). There was no age difference in average JOL. This could be due to self-terminated encoding (i.e., older adults could compensate for lower encoding efficiency by studying longer) or, perhaps, mid-point anchoring on the JOL scale by both age groups (representative of a similar degree of uncertainty regarding memory quality). Unguided participants were approximately 8% more confident following study, consistent with mean-level effects discussed earlier (i.e., that guidance lowered confidence by constraining encoding). The decomposition of the age interaction with strategy revealed that using a distinctive encoding strategy raised confidence by approximately 22% for younger adults and 9% for older adults, reflecting a weaker recognition of the benefits of distinctive encoding on the part of older adults at the item-level prior to the appreciation of any memory performance-related feedback experienced during phase 1 test. The decomposition of the condition interaction with strategy revealed that using a distinctive encoding strategy raised confidence by approximately 9% in the guided condition but only 4% in the unguided condition, again consistent with mean-level effects. For each letter increase in memory aid word length, confidence increased by about 0.5%. During response coding, an informal pattern was noticed by experimenters that longer memory aid words seemed to be related to more complex aspects of the stimuli being studied; it was for this reason that word length was included in this regression. The reliability of this effect offers indirect evidence that more complex encoding (which is likely more difficult and time-consuming) may have led participants to be more confident in their memories. However, when this variable was omitted from the model, the other substantial effects did not change materially. While word length may have accounted for additional variance in Phase 1 JOL, it did not appear to interact with other variables in the model (e.g., guided participants might have been more likely to think about their study strategy and produce more complex mediators than unguided participants, but this was not the case).

For each second increase in category generation RT, confidence decreased by about 0.07%, and for each second increase in memory aid generation RT, confidence decreased 0.3% for distinctively-encoded items and 0.09% for relationally-encoded items. In other words, as encoding fluency decreased, so did participants' confidence in

Table 3.1

Multi-level regression predicting phase 1 judgment of learning

Effect	Estimate	*SE*	*df*	*t*
Intercept	55.84	3.23	115	17.29***
Age (Younger Adults)	-4.36	3.76	115	0.25
Condition (Unguided)	7.67	3.82	115	2.01*
Strategy (Distinctive)	11.59	1.13	97	10.27***
Age (Younger) x Strategy (Distinctive)	8.94	1.32	97	6.79***
Condition (Unguided) x Strategy (Distinctive)	-9.77	1.46	97	-6.70***
Memory Aid Word Length (Within Subject)	0.47	0.13	3654	3.67**
Category Label RT (Within Subject)	-0.07	0.03	3654	-2.67**
Memory Aid Word RT (Within Subject)	-0.10	0.02	3654	-4.20***
Memory Aid Word RT (Within Subject) x Strategy (Distinctive Encoding)	-0.22	0.03	3654	-6.49***

Note. Abbreviations: RT = response time; *$p < .05$, **$p < .01$, ***$p < .001$

their ability to remember items later during testing. As discussed above, this relationship is often unwarranted (e.g., thinking of memory aid words quickly does not mean that related study targets items will be remembered well).

To summarize the findings regarding phase 1 JOLs and encoding fluency, many variables influenced JOL formation beyond fluency, which actually has relatively little impact on the memory confidence associated with a studied item. It would be of interest in the future to assess whether perceived fluency would have a larger impact on measured confidence, assuming a metacognitive perspective on learning and self-regulation of study behaviors (e.g., Hines et al., 2009; Robinson et al., 2006).

3.2 Free Recall Performance Measures

Given that the guided and unguided conditions only differed experimentally for participants in phase 1 of the task, data pertaining to free recall performance were analyzed separately for each phase, thereby treating phase 2 as a sort of transfer task. One might, for example, expect a difference in recall performance between unguided and guided participants in phase 1 of the task due to different encoding parameters, while any such differences might be changed during phase 2 when self-chosen encoding strategies could be implemented by all participants. Analyzing data for each phase separately prevented inappropriate comparisons of recall performance between phases for each experimental condition while allowing for an examination of recall differences due to age and encoding strategy. This caution was not necessary for the JOL analysis reported above because it was actually of interest whether or not confidence varied between phases as a consequence of receiving experimenter guidance in phase 1. For example, it was possible to investigate whether or not prior guided experience using of the relational strategy (which was used more frequently by guided participants in phase 1) affected the degree to which participants were confident in its use in phase 2. This confidence in each strategy can be assessed at the item level both in terms of their frequency of use and also the average level of confidence associated with their use at the item and global levels.

3.2.1 Uncorrected Free Recall

In terms of the raw proportion of studied items that were recalled at test, expectations were met in that distinctive encoding (M = 0.38, SE = 0.02) led to better memory outcomes than did relational encoding (M = 0.31, SE = 0.03), $F(1,101) = 7.88$, p

= .006, *d** = 0.21, during phase 1. Interestingly, younger (*M* = 0.35, *SE* = 0.02) and older adults' (*M* = 0.34, *SE* = 0.02) recall accuracy did not differ, $F(1,113) = .28$, $p = .597$. This equivalence was found despite younger adults' higher frequency of distinctive encoding. Perhaps it was the case that older adults were able to use purely relational strategies to generate high-quality mediators (i.e., they used their study time effectively), while that was not the case for younger adults.

The Age x Condition x Encoding Strategy interaction was reliable, $F(1,101) = 8.83$, $p = .004$ (see Figure 3.6). Guided younger adults' memory outcomes did not differ between strategies, whereas unguided younger adults used the distinctive encoding strategy more effectively than the relational strategy. This difference may have resulted from the use of relational encoding following a failed attempt at distinctive encoding, as some post-experiment reports indicated (i.e., younger adults in the unguided condition often reported using the relational strategy when distinctive encoding did not yield a high-quality memory aid word). Older adults, on the other hand, exhibited the opposite effect. When unguided, their memory outcomes did not differ between strategies, whereas under guidance, items encoded relationally were not recalled as accurately as items that were encoded distinctively. Younger adults performed best when selecting their own strategies (remember that the distinctive strategy was spontaneously used approximately 90% of the time, boosting absolute levels of recall), whereas older adults required explicit guidance toward distinctive encoding to boost recall performance for those items. Perhaps guidance led older adults to focus more purposefully on the distinctive properties of target items, which would have helped to differentiate target items from non-targets during testing.

In phase 2, those who had been guided strategically in phase 1 (*M* = 0.39, *SE* = 0.03) recalled more items correctly than those who had not been guided (*M* = 0.30,

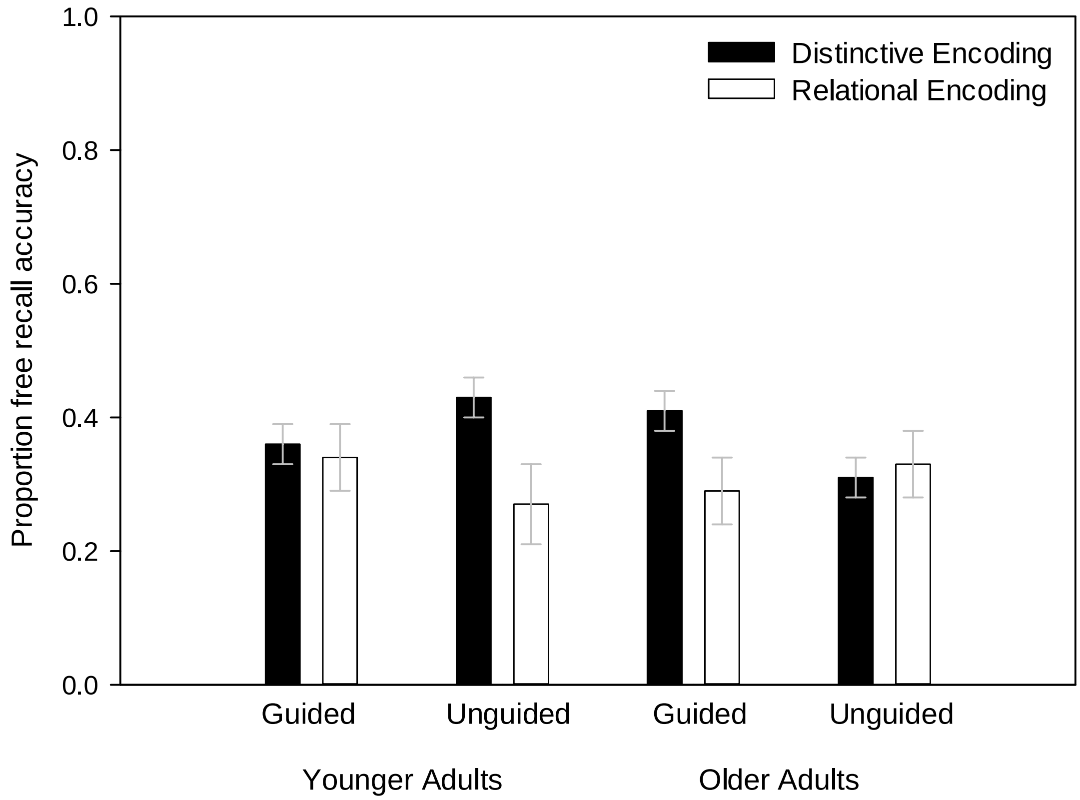

Figure 3.6. Proportion free recall in phase 1 by age, condition, and encoding strategy (means and standard errors)

SE = 0.02), *F*(1,92.8) = 6.63, *p* = .012, *d** = 0.53. A Condition x Strategy interaction was also reliable, *F*(1,83.7) = 8.11, *p* = .006 (see Figure 3.7). Distinctive and relational encoding led to similar recall accuracy for those who had been guided in phase 1, whereas those who were unguided previously displayed a clear disadvantage for items encoded with a relational strategy. The latter effect was clearly driven by younger adults, though the Age x Condition x Strategy interaction was not reliable.(*p* > .4) This finding suggests that guided strategy use in phase 1 might have led participants (or at least younger adults) to select a strategy for each list more appropriately in phase 2 (i.e., experience-based insight into appropriate strategy selection might have been more likely to develop following guided than unguided practice). On the other hand, unguided participants may have been at a disadvantage in remembering relationally-encoded items as a consequence of typically using the relational strategy when they were unable to generate a high-quality distinctive memory aid (thereby tying this strategy to items that participants thought were less likely to remember). The latter explanation is consistent with mean-level JOLs, which were approximately 10% lower for relationally-than distinctively-encoded items.

3.2.2 Intrusion Errors

The current task presented participants with many possible sources of intrusion errors, or items reported to be studied targets during a memory test, but which were actually derived from other sources. These intrusions could have resulted from confusion between studied targets and non-target list items, participant-generated categories or memory aid words, non-list words that were semantically-related to studied lists, or even practice items that were mistakenly classified as relevant studied targets during recall. It was therefore of primary importance in this task that participants distinguish between studied target items and non-target items in memory using any episodic contextual information about these items that existed in memory during testing to either accept a correct target or reject an intrusive item. Older adults commonly report more intrusions than do younger adults in free recall and recognition tasks (e.g., Murphy et al., 2007; Roediger & McDermott, 1995)-- a vulnerability that may result from frontal lobe deficits (e.g., Schacter et al., 1996; 1997).

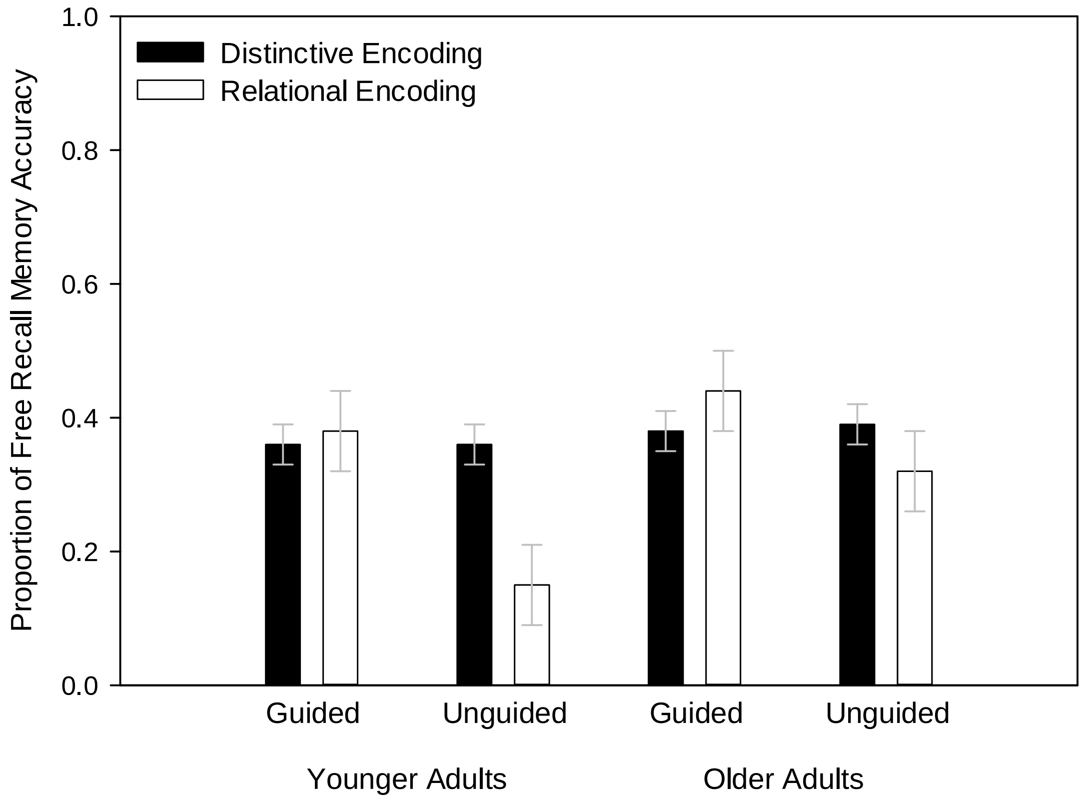

Figure 3.7. Proportion free recall in phase 2 by condition and encoding strategy (means and standard errors)

Intrusion errors were used to calculate the corrected free recall measure discussed in the next section. Prior to that discussion, a brief overview of the rates of intrusion from various sources is warranted due to the large number of sources from which they emanated and the fact that the intrusions were not limited to a few highly-errant participants but were instead distributed among the participant sample. See Table 3.2 and Table 3.3 for the distribution of intrusion errors in phases 1 and 2, respectively.

Two analyses are reported for each task phase. The first includes all intrusion errors, irrespective of associated study strategy (strategies were not recorded for practice lists and were not discernible for reported items that appeared to be

Table 3.2

Phase 1 intrusion errors by age, condition, and type of error

	Younger Adults		Older Adults	
Type of Response	Guided	Unguided	Guided	Unguided
Practice Target	49	52	27	30
Practice Non-Target	1	1	1	2
Practice Category	0	0	0	1
Non-Target List Item	4	5	11	18
Generated Word	0	1	4	9
Category Label	1	5	12	9
Related Non-List Item	5	0	4	7
All Intrusions	60	64	59	76
Accurate Responses	369	378	359	282

semantically unrelated to any studied lists). The second analysis includes only items within each task phase, thereby excluding items that were presented either in practice or a prior task phase. The latter analysis is of greater importance to the current research because it affords an opportunity examine the types of errors that followed each strategy. Rather than focusing on items that were, for example, learned so well in phase 1 that participants reported them again (errantly) in phase 2, it is of greater interest currently to ask questions such as whether or not participants reported non-target items more frequently following relational study.

Table 3.3

Phase 2 intrusion errors by age, condition, and type of error

	Younger Adults		Older Adults	
Type of Response	Guided	Unguided	Guided	Unguided
Practice Target	3	0	2	4
Practice Non-Target	0	1	1	2
Non-Target List item	1	1	2	6
Generated Word	0	0	8	5
Category Label	0	0	3	2
Related Non-List Item	0	0	3	4
Unrelated Word	0	0	1	0
Phase 1 Target	63	37	64	72
Phase 1 Non-Target	1	0	0	3
Phase 1 Generated Word	0	0	4	1
Phase 1 Category Label	0	0	0	1
All Intrusions	68	39	88	100
Accurate Responses	370	312	356	327

As can be seen in the data, the vast majority of intrusion errors were target items, either from practice trials (phase 1) or from a prior task phase (phase 2). Most of the intrusions therefore did not result from a failure to differentiate between target and non-target items in memory, but from a failure to limit reported items to a particular subset of items in memory (i.e., targets that were presented in during the most recent study phase). Interestingly, in phase 1, older adults reported fewer practice targets; this may have resulted from a higher (i.e., more conservative) response confidence threshold on the part of older adults. In phase 2, the majority of intrusion errors were targets from phase 1 that were encoded well enough to be recalled despite the interference created during the second study phase, and the number of intrusions were similar for younger an older adults. This increase in the number of intrusions for older adults may simply be an artifact of the larger number of potential intrusive items available in memory (i.e., there were only 5 practice lists but 32 phase 1 lists presented to participants). Even with a possibly higher response threshold, older adults may have experienced so much retrieval interference in phase 2 that intrusions became more frequent not only for phase 1 study targets but also for other stimuli present during study in both phases. However, as stated above, the current research is not focused on the differentiation of items based upon source (e.g., phase 1 or 2); instead, the current focus is on whether or not distinctive encoding afforded better target retrieval than relational encoding.

In terms of overall rates of intrusion errors during phase 1, unguided participants (M = 0.19, SE = 0.02) committed marginally more errors than did guided participants (M = 0.14, SE = 0.02), $F(1,114) = 3.66$, $p = .058$, $d^* = 0.35$. Study guidance might have afforded higher-quality encoding that was somewhat protective against intrusion errors. When study strategy was added to the ANOVA (and intrusions from non-study lists were removed from consideration), a clear age difference appeared with respect to within-phase intrusions, $F(1,190) = 12.04$, $p < .001$, $d^* = 0.58$. Younger adults (M = 0.03, SE = 0.02) exhibited a smaller proportion of errors than did older adults (M = 0.12, SE = 0.02). The Age x Condition x Strategy interaction was not reliable, $F(1,190) = .60$, $p = .441$, nor were the differences between distinctive and relational strategies within each age group, but a few mean differences are worthy of note that are otherwise obscured by a large amount of variability (See Figure 3.8). Younger adults' rates of within-phase intrusions

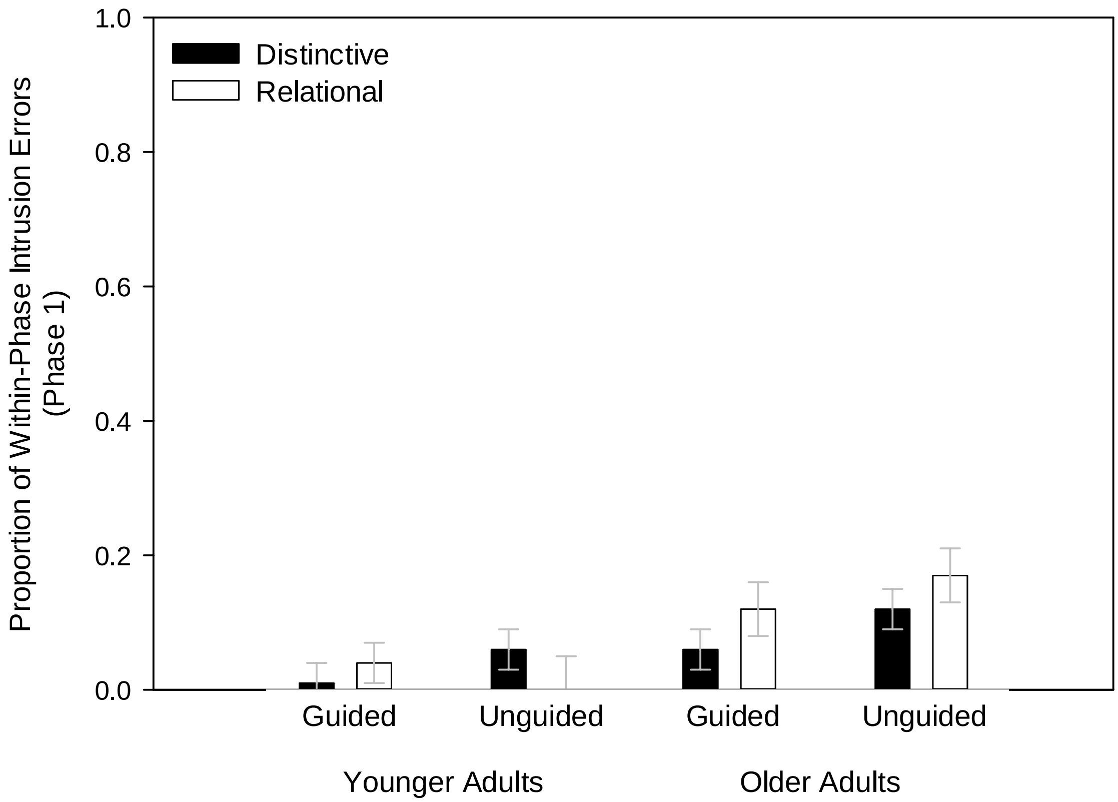

Figure 3.8. Proportion of within-phase intrusion errors for phase 1 by age, condition, and study strategy (means and standard errors)

did not differ from zero (all $p > .09$). Older adults were more vulnerable to intrusions--especially for relationally-encoded items. Unguided older adults also committed approximately 30% more intrusion errors than did guided older adults, indicating that guidance toward the use of specific strategies aided their implementation (e.g., by highlighting the unique properties of to-be-remembered items or, more generally, by providing more context during recall).

In phase 2, the overall rate of intrusion errors was larger for older (M = 0.21, SE = 0.02) than younger adults (M = 0.13, SE = 0.02), $F(1,114) = 15.58$, $p < .001$, $d^* = 0.39$. An Age x Condition interaction reflected a larger rate of errors for guided than unguided

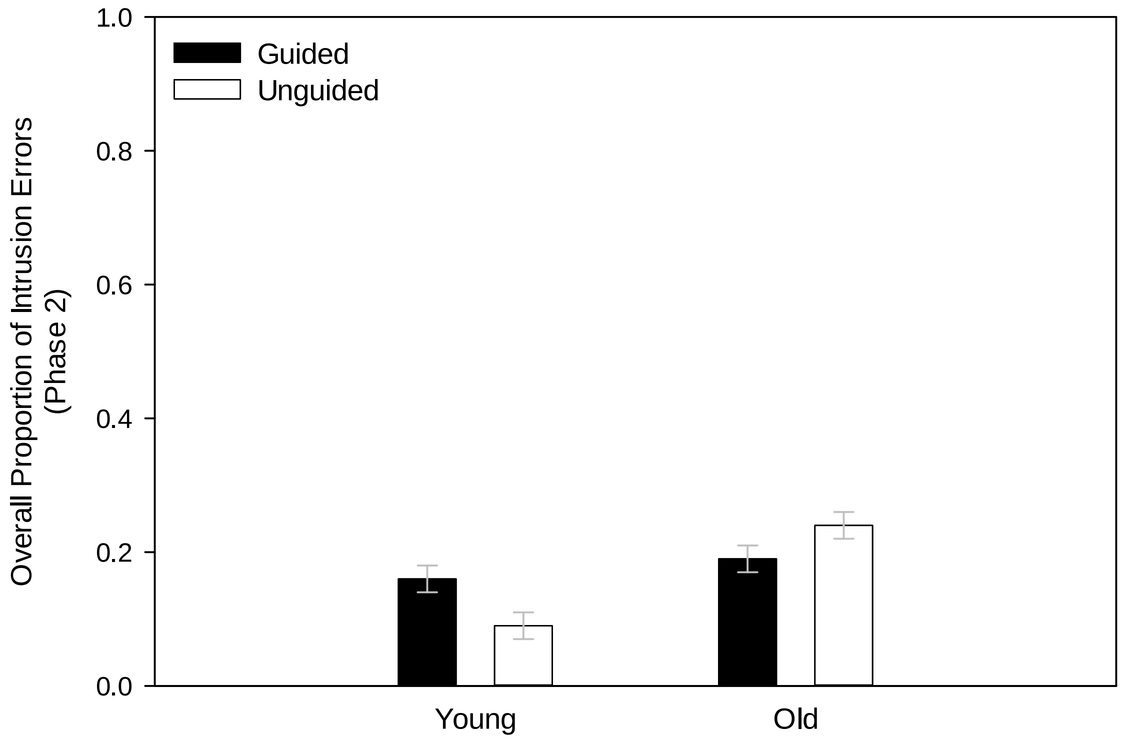

Figure 3.9. Proportion of all intrusion errors for phase 2 by age, condition, and study strategy (means and standard errors)

younger adults, whereas the rates did not differ between older adult groups (see Figure 3.9). Unguided younger adults simply reported fewer items from phase 1 than did the other groups.

An ANOVA focused on intrusions from phase 1 reported during phase 2 recall revealed that such intrusions were much more common for relationally-encoded (M = 0.35, SE = 0.03) than distinctively-encoded items (M = 0.13, SE = 0.03), F(1,181) = 27.30, $p < .001$, d^* = 0.82. Furthermore, the Age x Strategy interaction was not statistically significant, F(1,181) = 3.31, p = .070, though a trend did exist such that younger adults were actually more susceptible to intrusions for relationally-encoded items than were older adults, despite similar levels for distinctively-encoded items (see

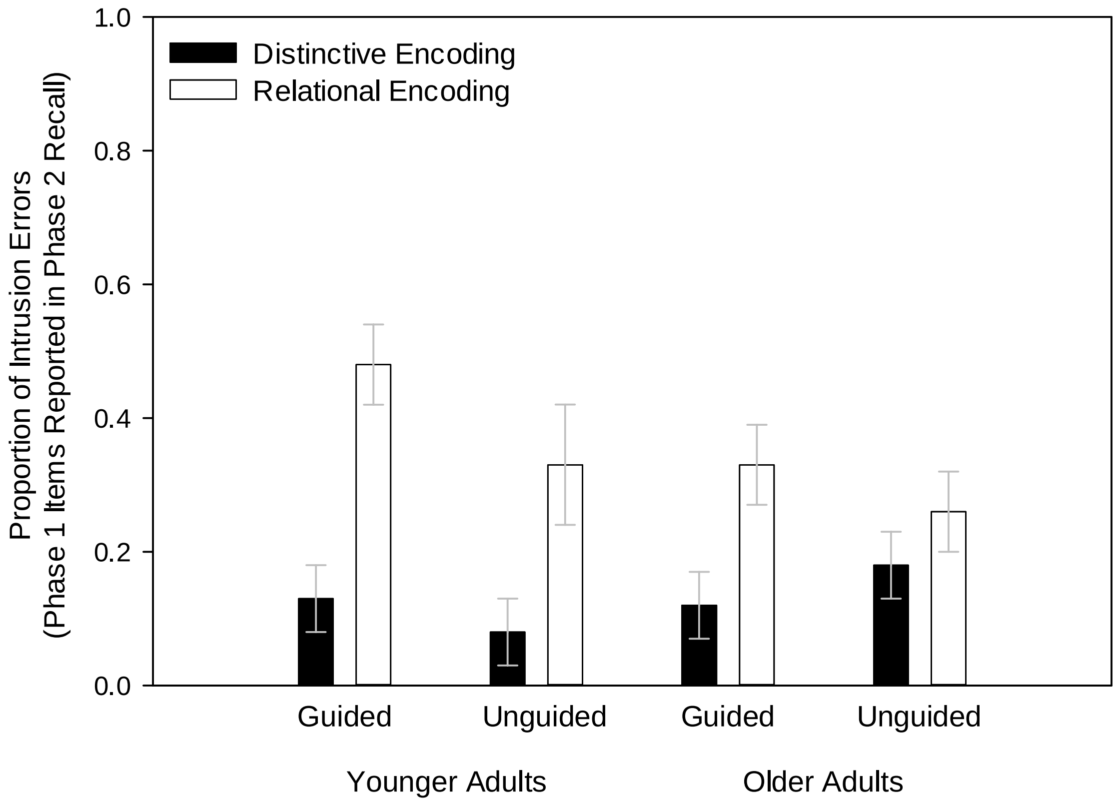

Figure 3.10. Proportion of all phase 1 intrusion errors in phase 2 by age, condition, and study strategy (means and standard errors)

Figure 3.10; each bar represents the number of phase 1 items errantly reported in phase 2 recall divided by the total number of items reported during phase 2 recall). Individuals can be trained to avoid such errors (e.g., Jennings et al., 2005), but in this task, they had very little to protect their recall from intrusions beyond effective distinctive encoding, which was not as beneficial for older adults as it was for younger adults in this regard.

Finally, an ANOVA focused on within-phase intrusions revealed that older adults (*M* = 0.04, *SE* = 0.01) again committed more errors than younger adults (*M* = 0.004, *SE* = 0.01), $F(1,163) = 11.55$, $p < .001$, $d^* = 0.55$. After removing items presented in phase 1 or during practice, the rate of intrusion errors declined considerably for both younger and older adults compared to phase 1. No other main effects or interactions were reliable.

3.2.3 Corrected Free Recall

Analyses pertaining to recall are often not corrected (unlike measures of recognition, which are often corrected by subtracting false alarm rates from hit rates; Dobbins et al., 1998). The current experiment, however, was created with the expectation that non-target items from the current task phase would be a primary source of intrusion errors in free recall, especially for relationally-encoded target items. With respect to phase 2 recall, for example, it was viewed as more important to examine variability in recall accuracy due to self-guided strategy use without the proportion of recalled items or intrusions being biased by additional items encoded under different conditions (referring back to the idea presented earlier that phase 2 can be viewed as a type of transfer task). The corrected free recall measure reported here is therefore the number of correct target retrievals for each participant minus the number of within-phase intrusion errors divided by the number of items studied for each strategy (intrusions from a prior phase were ignored in the current analysis).

In phase 1, items studied using the distinctive strategy ($M = 0.36$, $SE = 0.03$) were recalled at a higher rate than those studied with the relational strategy ($M = 0.29$, $SE = 0.03$), $F(1,101) = 6.30$, $p = .014$, $d^* = 0.25$. The Age x Condition x Strategy interaction was reliable, $F(1,101) = 5.64$, $p = .020$ (see Figure 3.11). Younger adults' corrected recall accuracy was similar for both strategies following guidance. On the other hand, younger adults who engaged in spontaneous strategy use performed much better with distinctive encoding than relational encoding. This finding may have resulted from unguided younger adults using relational encoding as a last-resort measure, whereas guided younger adults were encouraged to really try to use the relational strategy successfully. Alternatively, it may be possible that guided experience using each strategy changed how they were being implemented. For example, guided younger adults may have also used relational encoding as a last-resort measure, but the additional strategy-specific practice received during phase 1 may have increased the average quality of relational encoding beyond that which was afforded by spontaneous strategy use in phase 1.

Older adults' corrected recall accuracy was superior following guidance, but only for items encoded using the distinctive strategy. Experimenter-guided strategic implementation in phase 1 aided older adults' implementation of this strategy, possibly by making distinct item characteristics more salient. Unguided older adults exhibited similar corrected recall for both strategies, possibly because they viewed the relational strategy as being more viable and therefore may have invested more effort into encoding items in that way (to be discussed further). At the end of phase 2, participants were asked to describe their encoding experiences with each list (including strategies chosen and rationales for doing so). Reports of this additional requested information were infrequent, but do lend support to the notion that many participants (especially younger adults) did view relational encoding as a secondary method under conditions of self-guided strategic implementation. Older adults typically did not supply a rationale for their chosen encoding method, so it is harder to determine why they selected one strategy over another in most cases. The possible explanations for the results discussed in this section should therefore be examined more directly in future work.

In phase 2, older adults (M = 0.37, SE = 0.03) were marginally more accurate than were younger adults (M = 0.31, SE = 0.03), $F(1,93.8) = 2.52$, $p = .116$, $d^* = 0.32$. This pattern resulted from the low proportion of correct recall for unguided younger adults' relationally-encoded items. As speculated above, these items may have been more vulnerable to forgetting or interference as a consequence of lower-quality encoding (i.e., younger adults in some cases reported using relational encoding as a sort of last-

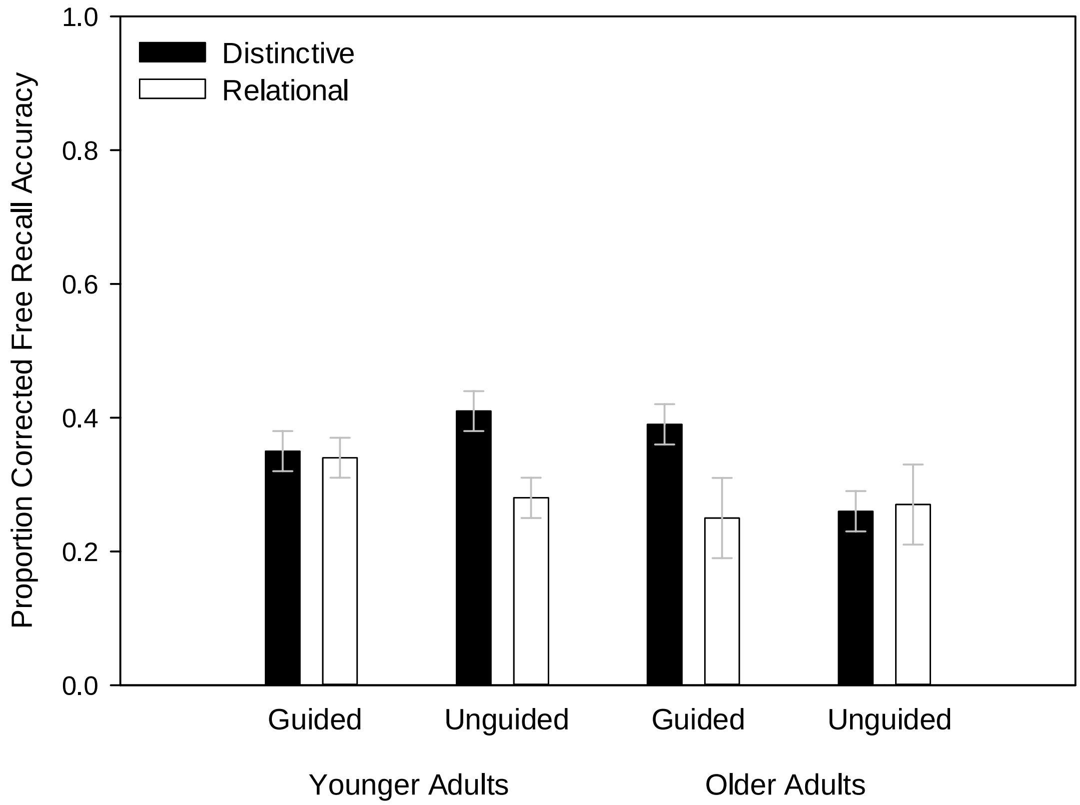

Figure 3.11 . Proportion corrected free recall in phase 1 by age, condition, and encoding strategy (means and standard errors)

ditch encoding effort). On the other hand, other groups (even unguided older adults) appear to have been more careful about the implementation of the relational strategy. Those who were guided in phase 1 (*M* = 0.38, *SE* = 0.03) were more accurate than those who were not (*M* = 0.30, *SE* = 0.03), *F*(1,93.8) = 5.53, *p* = .021, *d** = 0.34, lending support to the notion that experimenter guidance led to more competent strategy use in a general sense. Distinctive encoding led to marginally better recall accuracy (*M* = 0.37, *SE* = 0.01) than relational encoding (*M* = 0.31, *SE* = 0.03), *F*(1,83.4) = 2.27, *p* = .136, *d** = 0.33. This pattern was found only for those who were unguided in phase 1, *F*(1,83.4) = 6.46, *p* = .013, *d** = 0.73 (Distinctive: *M* = .37, *SE* = .03; Relational: *M* = .23, *SE* = .03;

see Figure 3.12). The Age x Condition x Strategy interaction was not reliable, p > .30. Guided participants (and unguided older adults, who, as will be discussed below, over-valued the relational strategy) evidently learned how to implement the relational strategy better due to more practice in its use during phase 1, whereas participants who seldom used the strategy (i.e., unguided younger adults) failed to use is as well in phase 2.

3.2.4 Free Recall Memory Confidence

3.2.4.1 Confidence for Accurate Responses

Younger adults (*M* = 97.04, *SE* = 1.85) were more confident in their free recall responses than were older adults (*M* = 88.95, *SE* = 1.82), *F*(1,126) = 9.73, *p* = .002, *d** = 0.55. The Age x Condition interaction was marginal, *F*(1,26) = 3.04, *p* = .084 (see Figure 3.13). Younger adults' confidence did not vary between guidance conditions, whereas older adults' confidence suffered in the context of unguided study (*M* = 85.36, *SE* = 2.63) relative to guided study (*M* = 92.55, *SE* = 2.53; *d** = 0.36), perhaps due to a correct perception of their lower degree of accuracy. As discussed previously, participants who engaged in guided study consistently studied longer than those who were unguided, and this may have also influenced memory confidence (e.g., items studied longer may have been more salient at test). Confidence also varied with study strategy, *F*(1,262) = 9.27, *p* = .003, *d** = 0.12. Distinctive encoding (*M* = 94.52, *SE* = 1.32) led to higher recall confidence than did relational encoding (M = 91.47, SE = 1.46), though both were near-ceiling. This is likely an artifact of test instructions indicating to participants that they should report only studied target items. Had a lower memory threshold been set by participants, a higher degree of variability likely would have been seen in memory confidence, as more items would likely have been reported.

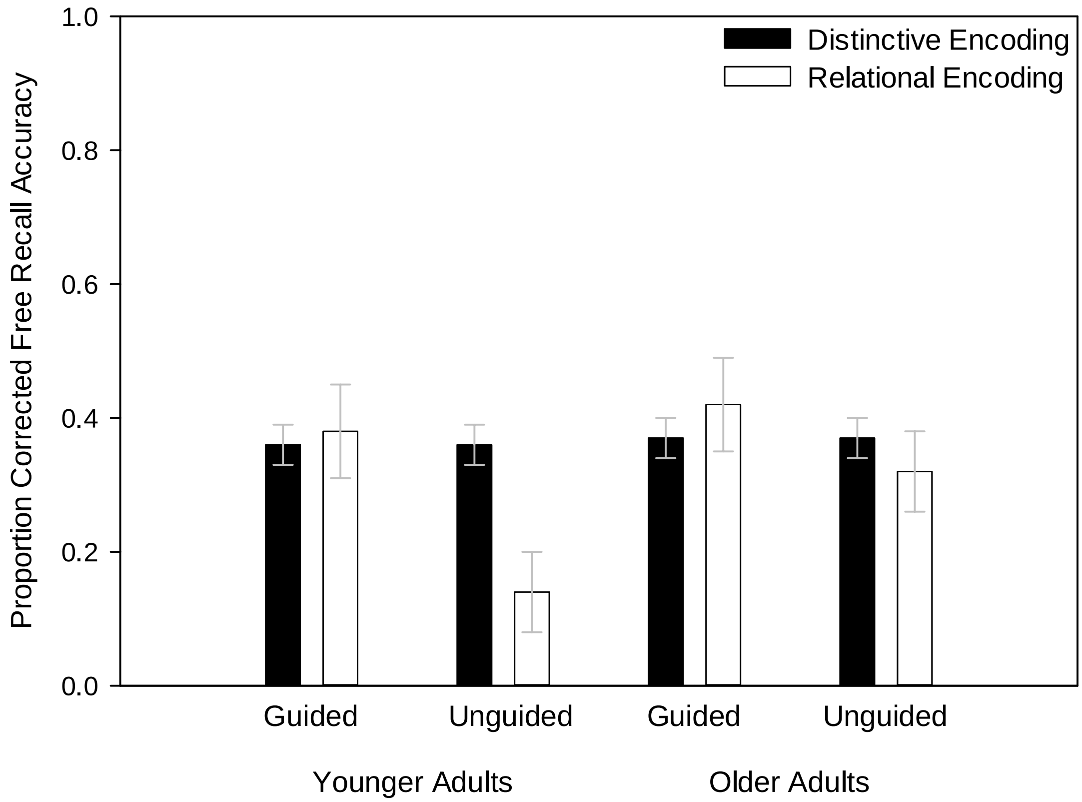

Figure 3.12 . Proportion corrected free recall in phase 2 by age, condition, and encoding strategy (means and standard errors)

3.2.4.2 Confidence for Intrusion Errors

The purpose of the following analysis was to examine the variability in confidence for memory errors. It may have been the case that the confidence associated with errant reporting was higher for older adults than younger adults, whose better episodic memories may have afforded a greater appreciation of the variability in response quality (i.e., younger adults may have been more aware that intrusion errors were lower-quality responses due to their greater episodic memory, but some responses still reached the required threshold for reporting). Confidence did not vary between sources of intrusion,

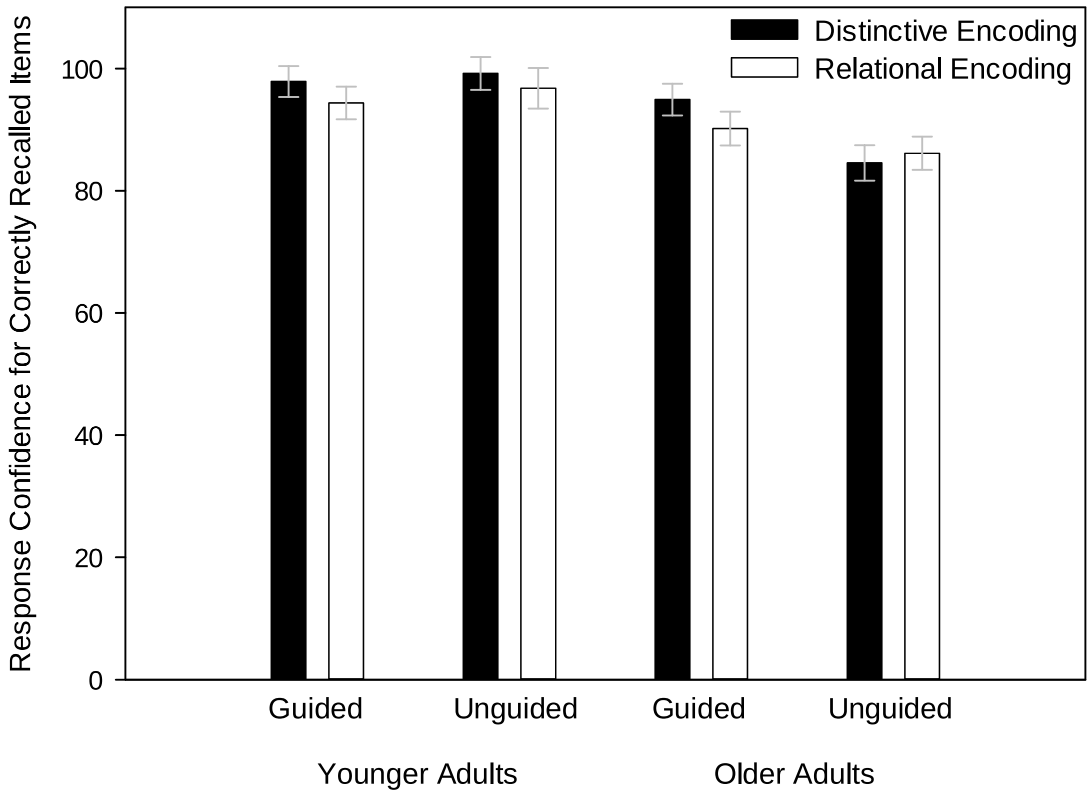

Figure 3.13 . Free recall confidence judgment for correctly recalled items by age, condition, and strategy (means and standard errors)

so all intrusion errors were included in this analysis. Contrary to initial assumptions, no age difference was found for confidence in intrusion errors, $F(1,112) = 1.64$, $p = .202$, $d^* = 0.11$. Younger ($M = 83.11$, $SE = 3.08$) and older adults' ($M = 77.51$, $SE = 3.09$) confidence in their intrusion errors was lower in magnitude (approximately 15% for younger adults and 8% for older adults) and more variable than was found for accurate responses. Confidence decreased between phase 1 ($M = 83.76$, $SE = 2.40$) and phase 2 ($M = 76.87$, $SE = 2.80$), $F(1,162) = 10.23$, $p = .001$, $d^* = 0.13$, perhaps indicating a lowered response accuracy threshold or a downgrading of confidence following implicit memory feedback during phase 1 testing.

3.2.4.3 Absolute Accuracy of Memory Confidence

Absolute accuracy was calculated at the item-level as accuracy (100 for an accurate response or zero for an intrusion error) minus the confidence judgment associated with that response (zero to 100 on a continuous scale), then averaged at the participant-level for the ANOVA. Positive values represent underconfidence, negative values represent overconfidence (perfect calibration is set at zero). As with the confidence levels reported above, absolute accuracy was examined separately for accurate and inaccurate responses.

With respect to accurate responses, younger adults (M = 4.03, SE = 1.89) were better calibrated than were older adults (M = 13.16, SE = 1.87), $F(1,119) = 6.74$, p = .011, d^* = 0.60. On the other hand, no age difference in calibration was found for inaccurate items (Young: M = -76.45, SE = 4.19; Old = M = -73.27, SE = 3.28), $F(1,125)$ = 0.36, p = .551. Older adults were underconfident in their accurate responses to a greater degree than were younger adults (indicating either worse item-level memory monitoring or simply less lower average memory confidence), but both age groups were extremely overconfident in their inaccurate responses.

3.2.4.4 Relative Accuracy of Memory Confidence

Gamma correlations (Nelson, 1984) were calculated to examine the relative accuracy (i.e., resolution) of subjective memory confidence to memory performance.[1] Correlations were computed between item-level judgments of learning made following each study trial and free recall accuracy to determine how well participants predicted

[1] Due to the high degree of recall confidence and low variability discussed previously, very few participants had calculable gammas between recall accuracy and confidence judgments, and many of those that were calculable were perfect correlations of either 1 or -1 (reflecting a lack of variability in response accuracy and confidence). Participants essentially reported an item when they were highly confident in their accuracy with few exceptions. Given this paucity of data with which to work, it was inappropriate to examine group differences relative accuracy of confidence judgments.

their future memory accuracy. Immediately following study, participants essentially had no idea whether or not they would remember a studied target later. Younger (*M* = 0.06, *SE* = 0.03; not different from zero, *p* = .089) fared no better than older adults (*M* = 0.0003, *SE* = 0.04; *p* = .992), *F*(1,109) = 1.39, *p* = .240. Gammas improved marginally from phase 1 (*M* = -0.01, *SE* = 0.04; *p* = .694) to phase 2 (*M* = 0.07, *SE* = 0.03; *p* = .022), *F*(1,108) = 3.77, *p* = .055, *d** = 0.11, but they were highly consistent with previous research showing the uncertainty under which immediate JOLs are constructed (e.g., Nelson & Dunlosky, 1991). These data may on the surface seem contradictory to that which was presented for mean-level JOLs (i.e., that distinctive encoding led to higher JOLs than did relational encoding). However, the multi-level regression analysis revealed that there were a number of influences on the formation of JOLs that may have increased the confidence ratings for distinctive encoding relative to relational encoding while not actually increasing the likelihood of retrieving a highly-rated items later (e.g., encoding fluency; Begg et al., 1989; Hertzog, et al., 2003).

3.3 Global Measures of Strategic Knowledge

Global measures of memory confidence and strategic effectiveness were collected before study and following test in each task phase to capture participants' confidence. These predictions and postdictions were typically not very well-calibrated in terms of actual memory accuracy (e.g., recall was littered with high-confidence memory errors). However, it is worthwhile to assess whether or not participants were able to form general heuristics regarding strategic effectiveness in the absence of highly-accurate item-level monitoring.

3.3.1 Pre-Study Measures

3.3.1.1 Strategy-Specific Confidence Predictions

For all strategy-specific global confidence and effectiveness predictions, full Age x Condition x Task Phase x Encoding Strategy ANOVAs were computed, but only main effects and interactions relevant to stated hypotheses are presented. Confidence was higher for the distinctive strategy (*M* = 62.18, *SE* = 1.68) than for the relational strategy (*M* = 50.96, *SE* = 1.65), $F(1,152) = 51.45$, $p < .001$, $d^* = 0.50$, indicating explicit knowledge of normative strategic differences in memory quality. This effect was driven primarily by differences in the younger adults, $F(1,152) = 15.67$, $p < .001$, $d^* = 0.54$ (Distinctive: *M* = 66.96, *SE* = 2.36; Relational: *M* = 49.55, *SE* = 2.31; see Figure 3.13), who perceived a greater benefit of distinctive encoding than did older adults. Confidence dropped between phases more so for the relational strategy (Phase 1: *M* = 56.23, *SE* = 1.76; Phase 2: *M* = 45.69, *S* = 2.09; $d^* = 0.39$) than the distinctive strategy (Phase 1: *M* = 65.46, *SE* = 1.80; Phase 2: *M* = 58.89, *SE* = 2.13; $d^* = 0.25$), $F(1,166) = 4.30$, $p = .040$ (see Figure 3.14), which reflected an appropriate larger downgrading for the less-effective strategy (in terms of overall recall accuracy and intrusions).

A reliable Age x Condition x Phase x Strategy interaction was found, $F(1,166) = 13.12$, $p < .001$ (see Figure 3.15). Younger adults initially expected the distinctive strategy to lead to better memory outcomes than the relational strategy prior to phase 1 study, whereas older adults did not report differentiated expectations. Following test experience in phase 1, guided younger adults downgraded their confidence in both strategies, appropriately reflecting lower-than-expected recall performance, but, somewhat surprisingly, maintained differentiated predictions despite similar levels of use and similar recall performance for each strategy. Guided older adults initially rated both strategies similarly, but they downgraded their confidence in phase 2 for the relational strategy, appropriately recognizing that it was associated with lower proportional recall performance. Unguided younger adults downgraded their ratings in both strategies slightly (but not significantly), reflecting appropriate initial recall expectations that did not

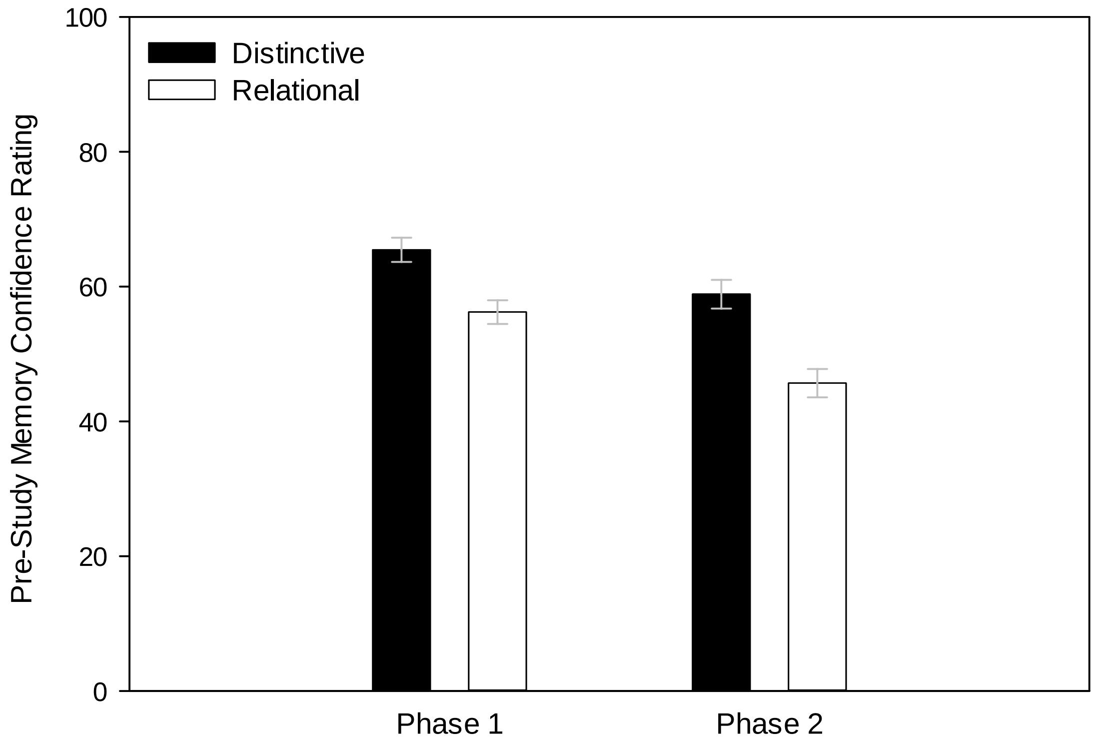

Figure 3.14. Strategy-specific global memory confidence prediction by age and study strategy (means and standard errors)

require much adjustment. Unguided older adults' memory expectations, on the other hand, were not altered following phase 1 test, appropriately reflecting similar levels of proportional memory accuracy for each strategy. Based on these data, it appears that both younger and older adults were able to track their recall performance in phase 1 and adjust their recall expectations in phase 2 appropriately.

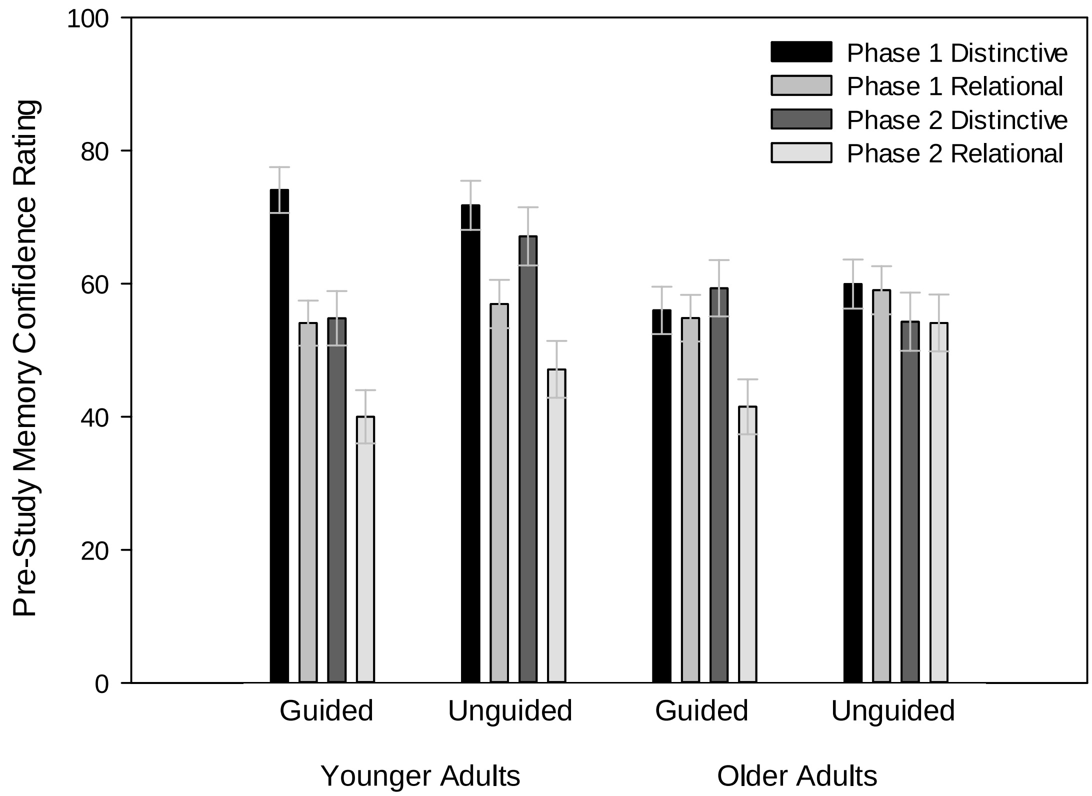

Figure 3.15. Strategy-specific global memory confidence prediction by age, condition, strategy, and task phase (means and standard errors)

3.3.1.2 Strategy-Specific Effectiveness Predictions

The distinctive strategy (*M* = 6.69, *SE* = 0.17) was rated higher in terms of effectiveness (on a scale from one to 10) than was the relational strategy (*M* = 5.47, *SE* = 0.15), *F*(1,146) = *p* < .001, *d** = 0.47. This effect was driven by younger adult ratings (Distinctive: *M* = 7.21, *SE*= 0.21; Relational: *M* = 5.16, *SE* = 0.21; *d** = 0.54; Older adults: Distinctive: *M* = 6.17, *SE* = 0.24, Relational: *M* = 5.79, *SE* = 0.21; *d** = 0.11), *F*(1,146) = 19.15, *p* < .001. The relational strategy (Phase 1: *M* = 5.95, *SE* = 0.16; Phase 2: *M* = 5.00, *SE* = 0.19; *d** = 0.39) declined in rated effectiveness between phases more so than the distinctive strategy (Phase 1: M = 6.92, SE = 0.18; Phase 2: M = 6.47, SE = 0.21; d^* = 0.17), $F(1,171) = 4.51$, $p = .035$.

The above effects were complicated by a reliable Age x Condition x Phase x Strategy interaction, $F(1,171) = 13.05$, $p < .001$ (see Figure 3.16). Similar to the global confidence data discussed above, both younger and older adults appear to have tracked their phase 1 recall performance fairly well. Guided younger adults downgraded their confidence in a general way, while still maintaining differentiated confidence. This may have resulted in pre-experimental strategic beliefs or their ratings may have been based on their perception of the number of items recalled rather than the proportion of items recalled for each strategy, given that the proportion of items was similar in phase 1. Unguided younger adults' ratings were also downgraded, but more so for the relational strategy than the distinctive strategy, which they correctly perceived as having led to better recall in terms of both the proportion of studied items and the absolute level of items recalled.

Guided older adults downgraded their confidence in the relational strategy, making use of implicit performance feedback from phase 1 test that was indicative of the lower quality of relational encoding. On the other hand, recall levels were similar for both strategies for unguided older adults, and this is reflected in the similar ratings for each strategy found in both phases of the task. These data, similar to that for confidence ratings, appear to reflect accurate strategy-specific memory monitoring in the current task, though it is possible that some ratings might also be influenced by strategic expectations derived from experience outside of the current task or even factors unrelated to the strategies themselves, such as a perception of poor strategic implementation (e.g., not taking enough time during study to think of high-quality distinctive memory aid words or setting a high recall threshold during testing).

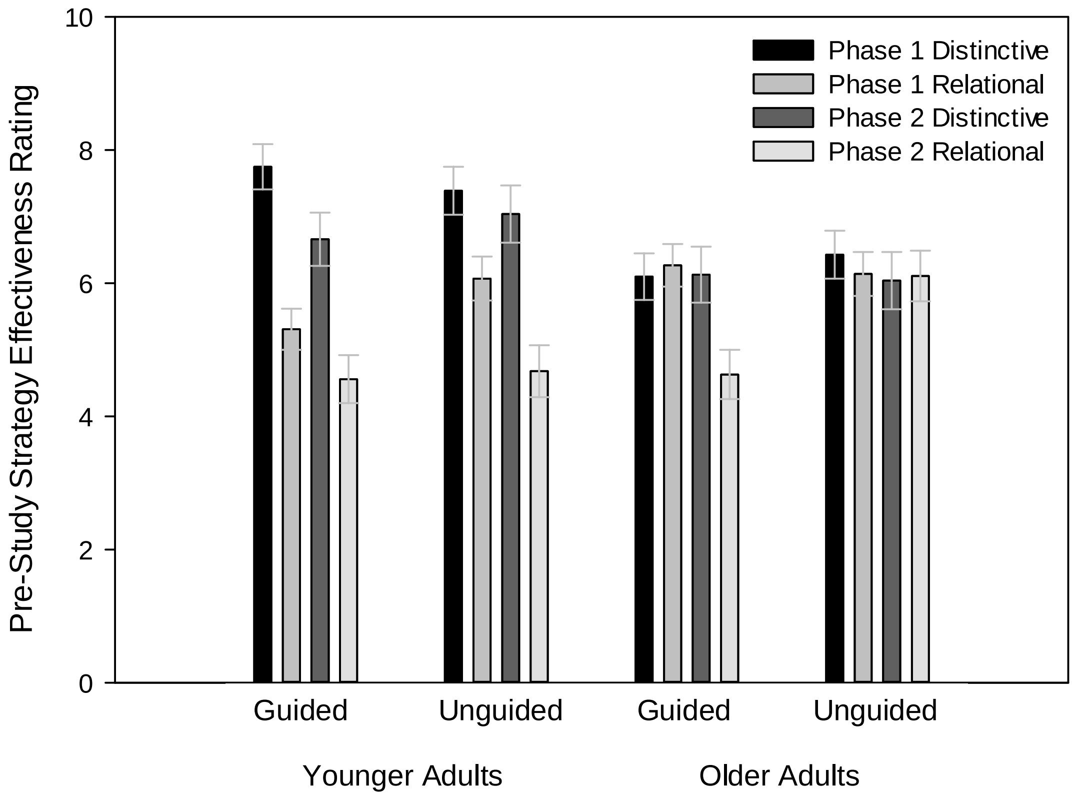

Figure 3.16. Global strategy-specific effectiveness prediction by age, condition, strategy, and task phase (means and standard errors)

3.3.1.3 Correlations Between Intrusion Errors and Strategy-Specific Predictions

To investigate the stark contrast between perceptions of strategy-specific memory performance and actual memory performance, Pearson correlations were computed between the proportion of intrusion errors and the difference between phase 1 and phase 2 reports of strategy-specific confidence and effectiveness. The rationale for this analysis was that if mean levels of recall accuracy (i.e., the proportion of correct responses) were not indicative of a differential benefit of memory (which is necessary for knowledge updating to occur, per Dunlosky & Hertzog, 2000), perhaps the differential perceptions of strategic success were tied instead to intrusion errors, which did vary somewhat between the different strategies. Unfortunately, it is hard to draw a supportive conclusion from the correlations, which did not near statistical significance (all p-values > .14). The highest correlations approximated .30, and reflected that as the number of phase 1 intrusions increased, associated global confidence predictions did as well (but, interestingly, not those related to effectiveness). With more data (i.e., a larger stimulus set or a larger number of participants), it might be the case that a more accurate estimate of these relationships could be assessed in a future study.

3.3.2 Post-Test Measures

3.3.2.1 Strategy-Specific Confidence Postdictions

Post-test global confidence ratings, which made use of recent memory performance feedback, are often considered one of the best indicators of strategy knowledge updating (see Hertzog et al., 2008). Confidence ratings were higher for the distinctive strategy (*M* = 43.68, *SE* = 2.32) than for the relational strategy (*M* = 30.48, *SE* = 2.05), *F*(1,154) = 36.49, *p* < .001, *d** = 0.44, and this difference was again driven almost completely by younger adults (Distinctive: *M* = 45.70, *SE* = 3.26; Relational: *M* = 25.21, *SE* = 2.88), *F*(1,154) = 11.13, *p* = .001, d* = 1.08 (see Figure 3.17). The difference between ratings for older adults was much smaller and not statistically significant (Distinctive: *M* = 41.67, *SE* = 3.31; Relational: *M* = 35.75, *SE* = 2.92), *t*(154) = 1.90, *p* = .060, *d** = 0.15. These ratings for older adults reflected actual memory performance in terms of the proportion of studied items that were recalled accurately at test, but failed to account for intrusion errors, to which they were highly vulnerable--especially for relationally-encoded items. They reported perceiving the relational strategy to be of similar quality to that of the distinctive strategy, which, unlike the relational strategy, was actually protective of memory errors in both younger and older adults. These results are aligned with those reported above for pre-study globally differentiated effectiveness ratings (Figure 3.16), indicating a real consistency in measures of

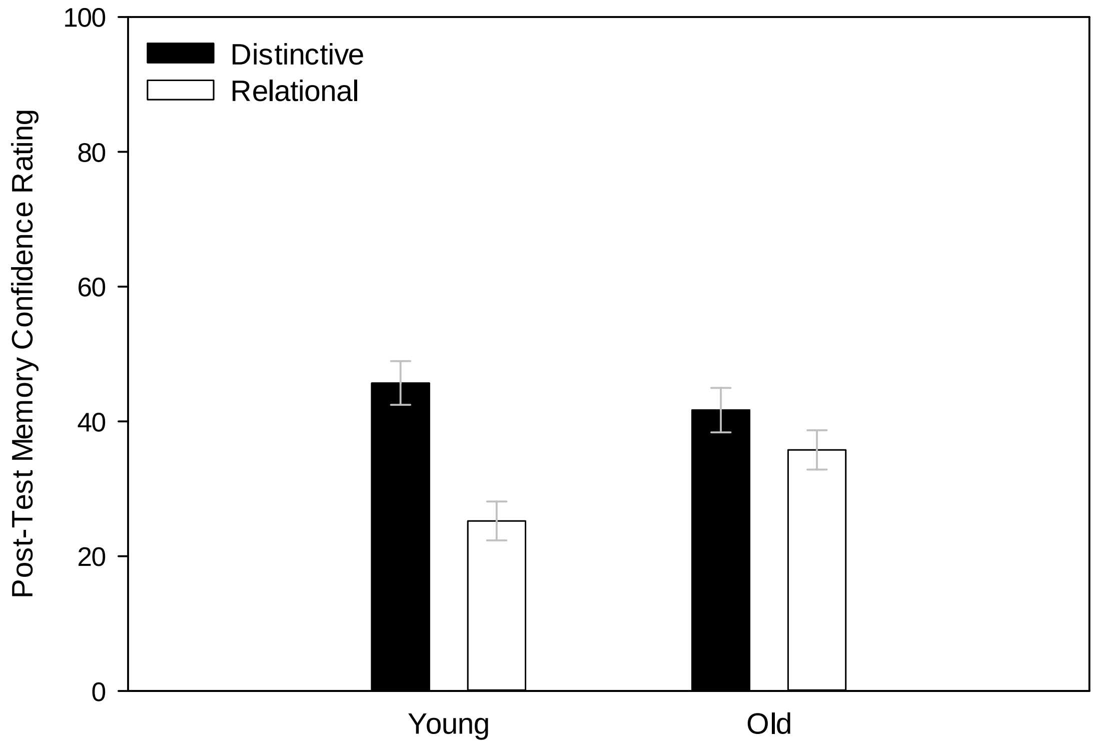

Figure 3.17. Global strategy-specific confidence postdiction by age, condition, and strategy (means and standard errors)

knowledge updating in this task. In terms of the proportion of items remembered for each strategy, both younger and older adults exhibit similar levels of global memory monitoring accuracy, but their monitoring did not appear to take into account intrusion errors, which would be expected given the relatively high confidence associated with such errors and the lack of objective memory performance feedback throughout the task that would have been necessary to highlight such errors.

A reliable Age x Condition x Phase x Strategy interaction was found for rated confidence, $F(1,160) = 13.73$, $p < .001$ (see Figure 3.18). Strategy-specific confidence remained largely unchanged for older adults regardless of condition, whereas that of younger adults fluctuated somewhat during the task. Guided younger adults' confidence

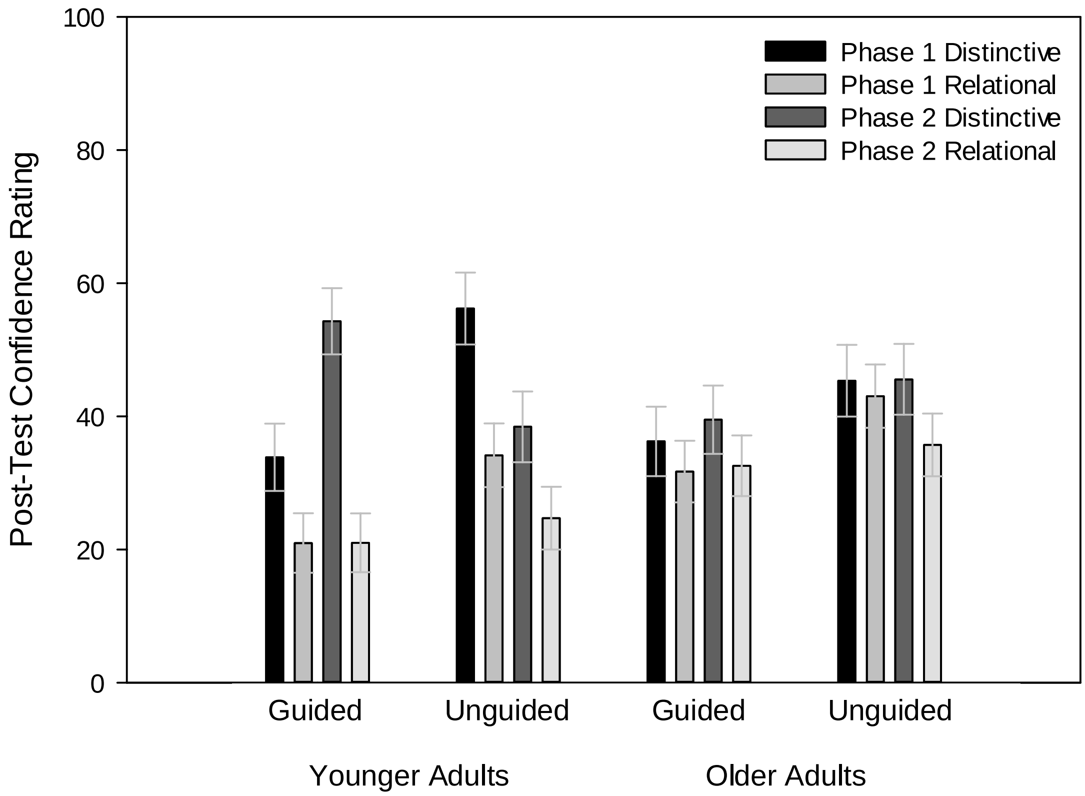

Figure 3.18 Global strategy-specific confidence postdiction by age, condition, strategy, and task phase (means and standard errors)

in the distinctive strategy increased greatly after the encoding restriction was lifted in phase 2, but confidence remained low for the relational strategy, appropriately reflecting its normative inferiority. Unguided younger adults' confidence in the distinctive strategy displayed the opposite pattern, declining sharply between phases—along with a smaller drop in confidence for the relational strategy. Confidence ratings did not change for older adults between phases, reflecting their more appropriate levels of initial confidence that did not require as much adjustment as did those of younger adults.

The absolute accuracy of post-test confidence ratings was calculated by subtracting a participant's post-test global differentiated confidence from their average level of corrected recall accuracy for each encoding strategy. Guided participants' postdictions (*M* = -36.11, *SE* = 2.58) were less overconfident than were those of unguided participants (*M* = -45.42, *SE* = 2.79), $F(1,160) = 6.01$, $p = .015$, $d^* = .43$. The Condition x Phase interaction was reliable, $F(1,118) = 10.72$, $p = .001$, as was the Age x Condition x Phase interaction, $F(1,118) = 8.47$, $p = .004$ (see Figure 3.19 for the latter interaction). Guided older adults were better calibrated than were unguided older adults in both phases, though neither group exhibited calibration changes between phases. Younger adults, on the other hand, were much better calibrated following strategic guidance in phase 1 than were those who were unguided. In phase 2, previously-guided younger adults were less calibrated than they were in phase 1, favoring expressing a degree of overconfidence similar to that of unguided younger adults. Though the precise mechanism is currently unknown, the strategic guidance that they experienced during encoding appears to have enhanced their ability to draw an appropriate inference regarding their recall performance. Perhaps guidance that increased younger adults' rate of relational encoding (that they know is of poor quality) led to a general downgrading of confidence that aligned more closely to actual performance. This might also explain why such an effect was not seen in the older adults, who initially valued relational encoding similarly to distinctive encoding.

The Strategy x Phase interaction was reliable, $F(1,149) = 6.32$, $p = .013$, as was the Age x Strategy x Phase interaction, $F(1,149) = 22.23$, $p < .001$ (see Figure 3.20 for the latter interaction). Younger and older adults' confidence postdictions were very similarly calibrated, except for those regarding younger adults' confidence in relationally-encoded items in phase 1. These data support the hypothesis above, which was that younger adults in phase 1 may have downgraded their confidence in the relational strategy relative to the distinctive strategy, which inadvertently caused their postdictions to align more closely to their actual memory performance.

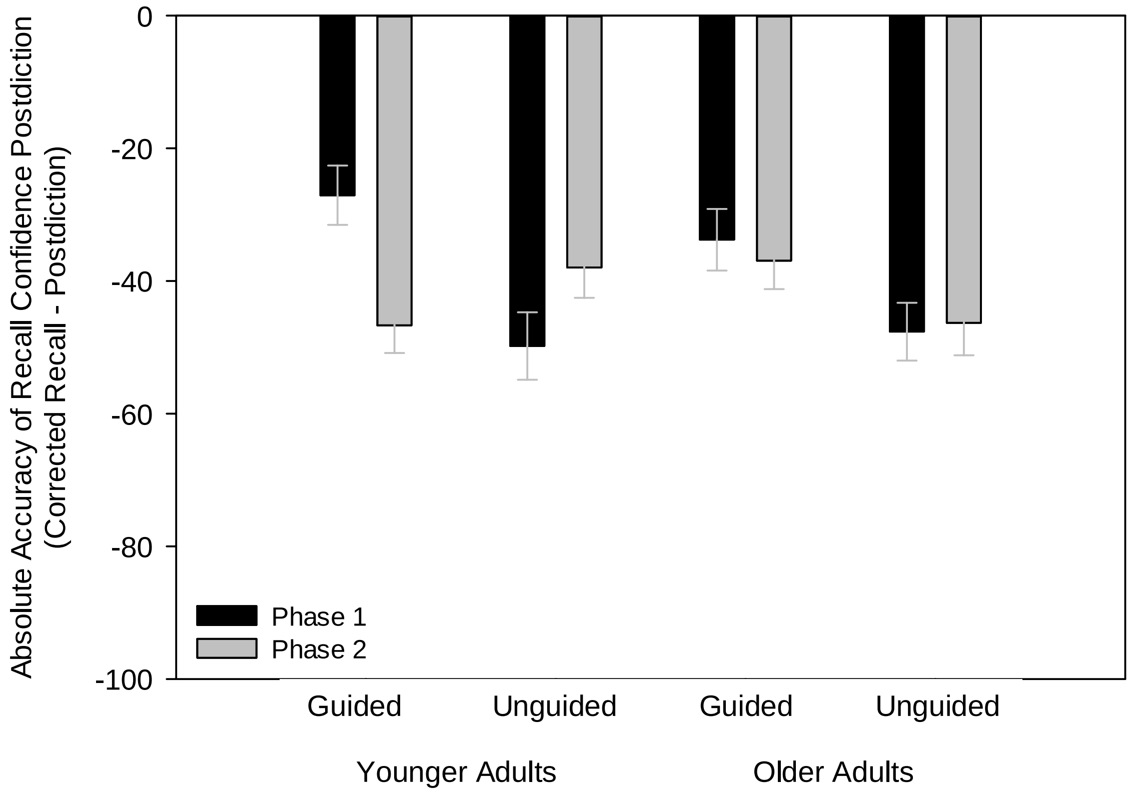

Figure 3.19. Absolute accuracy of recall accuracy postdictions by age, condition, and task phase (means and standard errors)

3.3.2.2 Strategy-Specific Effectiveness

Following testing, the distinctive strategy (*M* = 5.69, *SE* = 0.25) was rated as being more effective than the relational strategy (*M* = 4.38, *SE* = 0.21), $F(1,171) = 36.56$, $p < .001$, $d^* = 0.42$. This difference was driven primarily by younger adults (Distinctive: *M* = 6.07, *SE* = 0.35; Relational: *M* = 3.85, *SE* = 0.30; $d^* = 0.49$; Older adults: Distinctive: *M* = 5.31, *SE* = 0.36; Relational: *M* = 4.90, *SE* = 0.30; $d^* = 0.10$). A reliable Age x Condition x Phase x Strategy interaction was found, $F(1,139) = 4.04$, $p = .047$ (see Figure 3.21). Both younger and older adults' effectiveness ratings tracked actual memory performance, though it did not appear to factor in intrusion errors, which would have

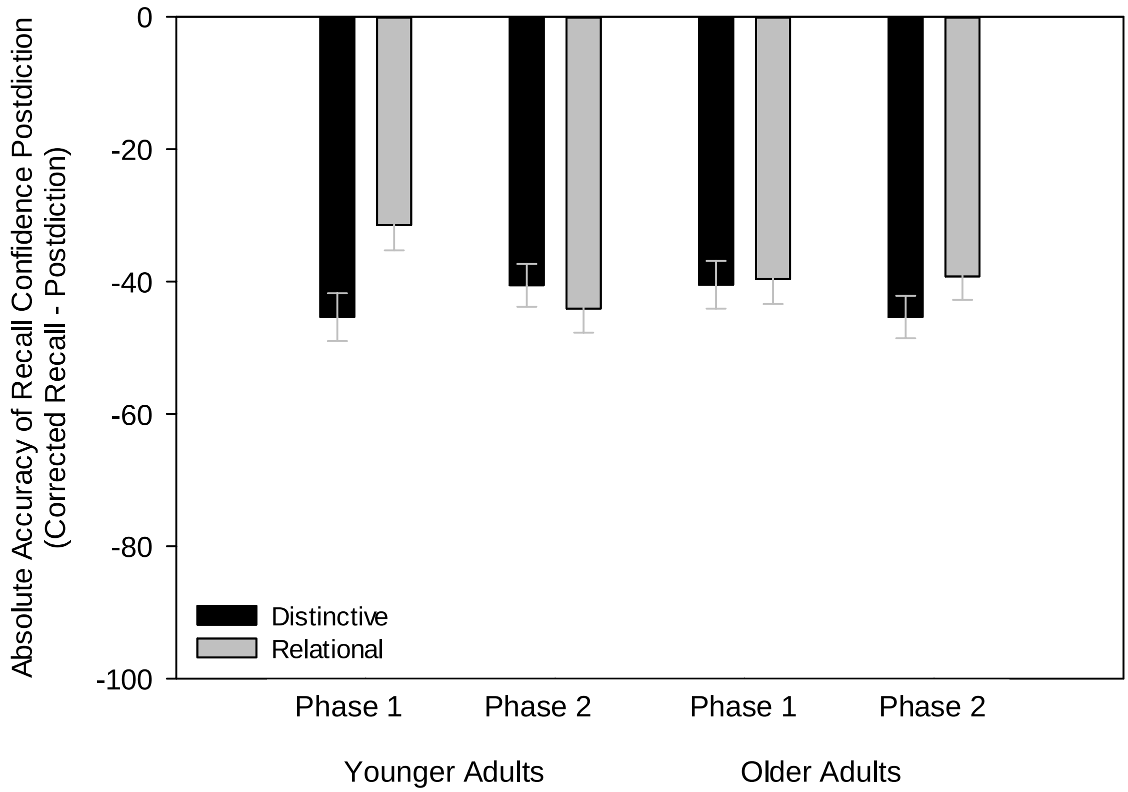

Figure 3.20. Absolute accuracy of recall memory confidence postdictions by age, strategy, and task phase (means and standard errors)

further differentiated the two strategies (especially for unguided older adults, who rated them most similarly throughout the task). As a final point, it also appears that younger adults' ratings--especially those of guided younger adults--were more separated than would be supported by the data alone, which is indicative of them capitalizing on their pre-existing knowledge of differential strategic effectiveness. Older adults appear to have either not used such normative strategic information or (as supported by pre-study ratings in phase 1) their pre-existing knowledge simply led them toward rating the strategies more similarly. Other evidence in the metacognitive literature supports the idea that older adults may also fail to track information pertaining to specific study

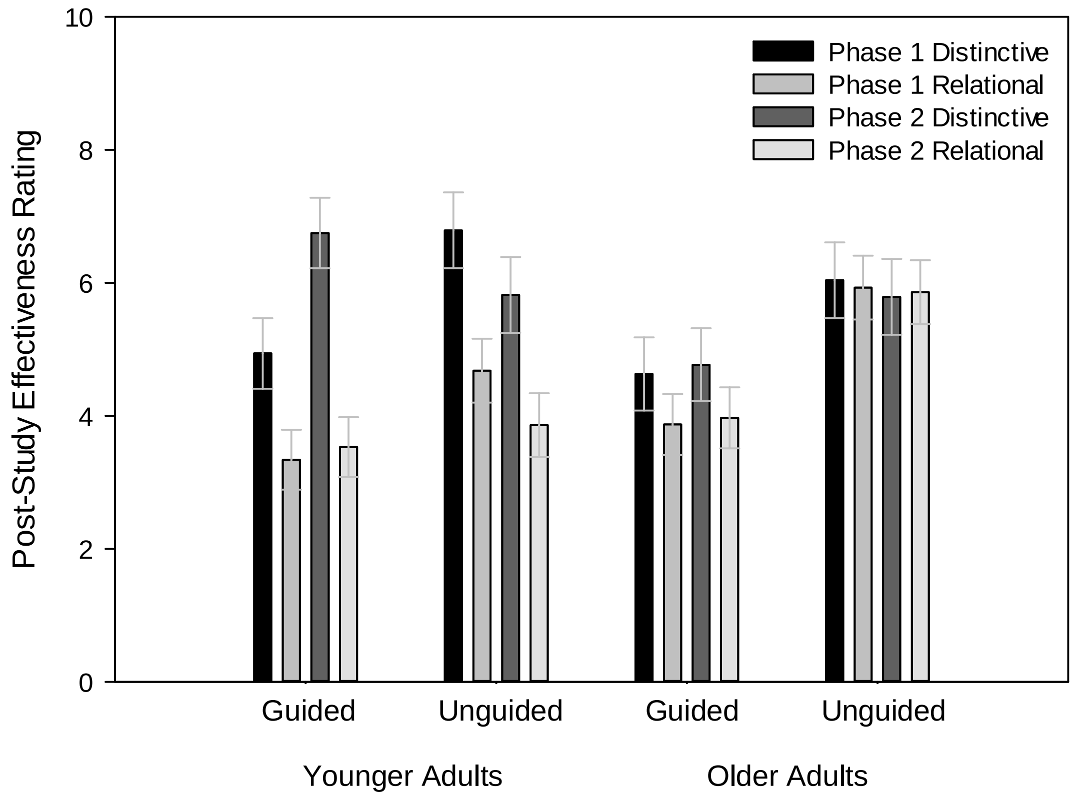

Figure 3.21. Post-test global differential effectiveness of encoding strategies by age, condition, strategy, and task phase (means and standard errors)

parameters (such as encoding method) over more lengthy retention intervals, which would certainly present a problem for knowledge updating (e.g., Tullis, Finley, & Benjamin, 2012).

CHAPTER 4

DISCUSSION

In the current study, it was hypothesized that both younger and older adults' memory performance would benefit from the use of a distinctive encoding strategy when compared to a purely relational encoding strategy. This hypothesis was generally unsupported: both younger and older adults generally appeared to implement both strategies with similar levels of success. Global measures of strategic knowledge indicated that younger, but not older adults, were explicitly aware of the normative superiority of a distinctive encoding process above that which was purely relational prior to any task experience, but they appeared to anticipate or perceive benefits associated with a distinctive encoding strategy that were not evident in their actual recall performance. The task afforded the opportunity for participants to learn more about the differential normative effectiveness of distinctive and relational encoding strategies, but, for most participants, this difference was not statistically significant. In phase 1 of the task, only guided younger adults and unguided older adults exhibited a difference in the expected direction, and in phase 2, this difference was found only for unguided younger adults.

Despite similar recall performance between strategies, an increased differentiation of strategic effectiveness ratings was seen in all groups except for older adults who did not experience experimenter-guided strategic implementation in phase 1 of the task. This effect was larger for younger adults, whose ratings became much more differentiated than they were initially, than for older adults, who tended to simply downgrade the less effective relational strategy following guidance. An explanation of this pattern of differentiation can potentially be found in data pertaining to memory intrusions present in free recall. Without any difference in recall performance between the strategies, it is possible that participants factored in these intrusions (which were much more common for relationally-encoded items) when thinking about strategic effectiveness. The current data were too variable to assess this question adequately, but this idea and others are discussed below in the context of cognitive aging research related to distinctiveness, strategy selection, and metacognitive knowledge regarding strategic effectiveness.

4.1 Effects of Distinctive Encoding

4.1.1 Free Recall Accuracy

Prior to any discussion related to metacognitive strategic knowledge, it is necessary to first discuss whether or not participants had the opportunity to perceive a difference in strategic effectiveness. Consistent with prior research (see Smith, 2006 for a detailed review), the combination of relational and item-specific encoding (i.e., distinctive encoding) led to a clear memory advantage over relational encoding alone. This advantage was present in many aspects of the data. In terms of uncorrected recall accuracy, the proportions of items recalled given a particular study strategy did not always differ for each experimental group, but the raw count of correct responses was always greater for the distinctive strategy. For example, in phase 1 unguided older adults recalled approximately 33% of items studied with each strategy, but they studied approximately 80% of all items using the distinctive strategy, leading to an actual ratio of recalled items of about 8 distinctive items to 2 relational items recalled per person. This is not a trivial difference, and participants therefore may have been able to detect it if they were, for example, attending to strategy information during recall and associating recall success rates with specific strategies.

The Contextual Support for Similarity and Difference (CCSD) framework (Smith, 2011) can be invoked to clarify the current findings regarding age similarities in distinctiveness effects. This framework was created to offer an explanation for the often mixed results regarding whether or not both younger and older adults can make use of the distinctive qualities of to-be-remembered information. Essentially, older adults can be expected to benefit from distinctive processing (at encoding or retrieval) if the task environment provides support that compensates for age-related declines in the resources necessary to either attend to, encode, or retrieve relevant item-specific information within a larger relational context. In the current task, such support is provided during encoding by requiring participants to first generate a relational context for each presented list (an explicit category label for the semantically-related items) and then request that participants engage in further encoding for a particular list member. This further encoding was sometimes explicitly requested (i.e., through guidance in phase 1), but even when it was not, support for distinctive encoding may have been provided simply as a consequence of informing participants about the existence of relevant encoding strategies (e.g., Dunlosky & Hertzog, 2001; Hertzog et al., 2012), which may lead to self-initiated strategic experimentation (e.g., Crowley, Shrager, & Siegler, 1997).

The CCSD framework (Smith, 2011) is framed in the context of findings related to age-related declines in cognitive resources (e.g., Verhaeghen, Marcoen, & Goossens, 1993) that may be compensated for through environmental support (e.g., Craik, 1986). A common conceptualization of these cognitive resources is working memory capacity, which, when reduced in quality by aging (e.g., through a reduction in cerebral blood flow or neurological degradation) or otherwise taxed as a consequence of engaging in a difficult task, may reduce the degree to which individuals are able to take advantage of distinctive information during encoding (e.g., Smith & Engle, 2011). Butler et al. (2010) proposed that focusing on item-specific information at encoding may cause a degradation in the quality of relational encoding in older adults (also see Hege & Dodson, 2004; McCabe & Smith, 2006), though the current task likely compensates for any such potential degradation through the explicit requirement of relational encoding for all items. Likewise, engaging in high-quality encoding might have a negative effect on retrieval processes by offering too much candidate information to sift through within an allotted timeframe, which could limit the retrieval of desired information. The latter consequence might account for the fact that, even given unlimited study time, neither older nor younger adults recalled more than 45% of items that were studied distinctively, even when given relatively small memory sets of, at most, 32 target items per phase. An alternative explanation for this level of recall would be that neither age group likely developed an optimized encoding or retrieval strategy for this novel task in the time available to them. However, distinctive encoding still led to better memory performance for both unguided younger and guided older adults, so, while some of these processes may have played a role in memory performance, distinctive encoding was still found to be somewhat effective in enabling subsequent recollection of studied items.

4.1.2 Intrusion Error Rates

Distinctive encoding also lowered the rates of within-phase intrusion errors and intrusions from a prior phase. Relational encoding therefore led to the recollection of phase 1 targets in many cases during phase 2 despite its normatively lower effectiveness, but participants appear to have been less capable of reinstating the study context to the degree necessary to realize that those items were from phase 1. In other words, relational encoding sometimes created item memories strong enough to persist into the next phase, but participants lacked enough contextual information to recognize their source. On the other hand, phase 1 items encoded distinctively were much more likely to be filtered from recall in phase 2 despite being recalled at greater rates during phase 1.

These results are consistent with prior research showing that distinctive encoding aids the prevention of false memories in the Deese/Roeidger-McDermott paradigm (DRM paradigm; Deese, 1959; Roediger & McDermott, 1995). Relevant to the current experiment, McCabe, Presmanes, Robertson, and Smith (2004) found that instructions to generate a unique feature of studied items on a list reduced younger adults' false memories for those lists. Butler et al. (2010) examined this phenomenon with older adults, but found that distinctive encoding actually interfered with older adults' memory performance (by decreasing accurate recall and increasing intrusion rates). The authors postulated that performance may have suffered for older adults because they either failed to perform distinctive encoding effectively within the allotted study time or because this additional information from encoding produced increased source interference at test. The relatively high rates of correct recall for distinctively-encoded items compared to relationally-encoded items in the current study, along with the fact that participants were allowed to terminate study themselves, suggests that older adults in this experiment were not limited with respect to their potential for a high degree of encoding quality. An age-related deficit in the retrieval of source information (e.g., Butler et al., 2004; Naveh-Benjamin & Craik, 1995; Zacks, Hasher, & Li, 2000), however, might explain why older adults were more vulnerable to intrusion errors in this task compared to younger adults--even for distinctively-encoded items, for which source information would likely have been available in memory. This argument is consistent with the findings of Gallo and colleagues (e.g., Gallo et al., 2007) showing older adults can utilize a distinctiveness heuristic to suppress false recognition of pictures compared to (less visually rich, or distinct) words, they were still at a disadvantage with respect to levels of source confusion for studied items compared to younger adults.

4.1.3 Effects of Strategic Guidance

Strategic guidance was implemented for the purpose of aiding strategic knowledge updating, however, it appears that explicit study instructions given at the trial-level aided some aspects of recall as well. Specifically, guidance may have afforded higher-quality encoding, which resulted in a lower rate of intrusion errors for older adults, who were generally more vulnerable to such errors. This effect was not limited to distinctively-encoded items; it appeared for relationally-encoded items as well, which indicates that older adults may have generally benefitted from a more structured (i.e., supportive) encoding environment--even when using a normatively less effective strategy (e.g., Smith, 2011).

4.2 Strategic Behavior

4.2.1 Strategic Implementation

Both younger and older adults displayed a clear preference for a distinctive encoding strategy over a purely relational strategy. When they were allowed to self-select strategies in either task phase, younger adults chose to use distinctive encoding on approximately 90% of study stimuli, and older adults chose to implement it approximately at a rate of approximately 80%. Contrary to expectations, the rates of distinctive encoding did not differ in phase 2 based upon whether or not participants experienced experimenter-guided study in phase 1. Instead, phase 1 guidance seems to have simply constrained participants' encoding, and after the constraint was removed, they resumed their normal patterns of use. This is highly indicative of pre-task knowledge of the superiority of the distinctive encoding strategy, but these findings may have also been biased as a consequence of informing participants about relevant study strategies and asking them to think explicitly about their effectiveness (e.g., Dunlosky & Hertzog, 2001; Hertzog et al., 2012). Researchers should consider implementing uninstructed phase 1 study to examine this possibility (e.g., as in Dunlosky & Hertzog, 2001). Additionally, the abnormally high reported usage of distinctive encoding on the part of older adults was not due to errant strategy reporting; participants were well-trained and observed throughout their participation. Also, there was a large temporal delay between actual encoding and strategic reporting, so the likelihood of participants actively engaging in deception seems low. If their memories were good enough to accomplish that, they should have performed better on the free recall task.

4.2.2 Adherence to Study Guidance in Phase 1

One might assume that, given participants' clear preference for distinctive encoding when they were permitted to self-select strategies, that a failure to adhere to study instructions would have usually resulted in the use of distinctive rather than relational encoding. However, as the data revealed, this was not the case. Though participants did defy experimenter guidance more frequently for distinctive than relational trials (60% of the instances of defiance were for guided distinctive trials), strategic adherence was quite similar for both strategies. Post-task strategy usage information (gathered during strategy reports) indicated that participants typically chose to defy guidance when they were unable to generate an adequate memory aid using the experimenter-requested strategy. Older adults may find it especially hard to generate distinctive features in the context of semantically-related lists (e.g., Butler et al., 2010). In the current task, participants young and old also reported difficulties generating a relational memory aid word following the generation of an also-relational category label--especially when the list members had only one common feature (e.g., SUPERMAN and an AIRPLANE both fly, but what else do they have in common?).

4.3 Strategic Knowledge

4.3.1 Item-Level Memory Confidence

Item-level judgments of learning at study were not predictive of future item recall for either younger or older adults. Item-level confidence judgments at test were likewise not very discriminative between instances of accurate target recall and intrusion errors, either. The correlation between recall accuracy and confidence was near zero, resulting from a small degree of variability in the data (i.e., participants responded only with items for which they held a high degree of confidence). However, confidence was approximately 10-15% higher on average for correct than incorrect responses, indicating that participants perceived a slight discrepancy between these responses, but were not necessarily willing to risk reporting items for which they were not very confident. The current task did not force participants to report studied items by, for example, using cued recall because pilot testing indicated that both recall accuracy and memory confidence measures would have been at ceiling for younger adults. These ceiling effects would have greatly limited the quality of the current examination of age differences in distinctiveness effects and strategic knowledge updating. Though item-level variability was still somewhat limited in the current work, the same was not true of global measures.

4.3.2 Global Differentiated Confidence and Effectiveness Ratings

Strategy-specific learning can be manifested in a number of ways, with some of the best evidence taking the form of modified (i.e., updated) strategy-specific effectiveness ratings, confidence postdictions following a test experience, and confidence predictions that precede a second (or subsequent) task phase (e.g., Hertzog et al., 2008). The precise mechanisms that afford knowledge updating, however, are not currently known. It is possible that there is a relatively automatic metacognitive monitoring process that stores memory retrieval outcomes (i.e., successes and failures) for later retrieval (e.g., White & Wixted, 1999). However, it is more likely that the products of related monitoring processes are not made available in a veridical, undistorted way to an individual. Instead, a more heuristic-based inferential process may be enacted that allows for somewhat accurate monitoring of retrieval outputs, but is imperfect (e.g., Brown, 1995; Manis et al., 1993; Winne, 1996). It is often the case, for example, that individuals have a sense that one strategy is superior to another, but they do not accurately estimate the differential effectiveness (e.g., Koriat & Bjork, 2006). However, even though the processes that may lead to knowledge updating are imperfect perceptions of performance, they are still necessary for knowledge updating to occur (e.g., Dunlosky & Hertzog, 2000; Finn & Metcalfe, 2007) and often correlate highly with strategic selection (e.g., Hertzog & Dunlosky, 2004; Rabinowitz, Freeman, & Cohen, 1992; Schunn & Reder, 2001).

The current data lend support to an inferential account of knowledge updating by showing that participants in many cases possessed explicit knowledge of the memory performance associated with each strategy, but their monitoring accuracy was far from perfect (e.g., post-test confidence was consistently overestimated by 30 to 45%). Both younger and older adults exhibited some degree of strategic knowledge or knowledge updating, when recall performance was supportive of it. Key findings of the current study are that older adults exhibited less knowledge updating than did younger adults, and participants who received strategic guidance in phase 1 generally exhibited more updating than those who did not. The latter finding indicates that explicitly associating encoding strategies with memory outcomes may help both younger and older adults perceive strategic differences.

In the absence of strategic guidance in the first phase of the task, older adults consistently failed to differentiate between the strategies in terms of associated confidence or effectiveness ratings. Contrary to expectations, this lack of observed change in strategic knowledge did not necessarily result from an age-related failure to monitor memory outcomes accurately; it resulted from the memory outcomes themselves being similar between the strategies (at least in terms of the proportion of items recalled). Older adults who were guided in the first phase demonstrated a pre-existing knowledge of differential effectiveness that was small (<10% difference between strategies), but this difference did not grow, except with regard to pre-study estimates of confidence and effectiveness (i.e., after phase 1 testing, they appeared to realize that relational encoding was less useful than initially thought). Guided older adults appropriately downgraded their expectations for the relational strategy, while holding those for the distinctive strategy constant. Unguided younger adults demonstrated a slight downgrading for both strategies (an approximate change of 15% for confidence and 10% for effectiveness ratings), correcting initial overconfidence and over-estimated effectiveness, but failing to further differentiate between the strategies. Younger adults who were initially guided demonstrated the greatest degree of knowledge updating of any group (e.g., the difference in post-test confidence increased from approximately 15% in phase 1 to 35% in phase 2 and that for effectiveness ratings increased from approximately 15% to 25%). This degree of change is not necessarily supported by raw proportions of item recall, but it may have been afforded by the explicit (and instructed) comparison of the two strategies.

A similar global downgrading was found in older adults by Bieman-Copland and Charness (1994) and Matvey et al. (2002) and is suggestive of an inferential deficit whereby participants can perceive components of their performance, but not necessarily draw global inferences about strategic effectiveness from those perceptions. Unlike a recent knowledge updating study involving experimenter- and self-chosen strategies for learning paired associates (Hertzog et al., 2012), the current evidence is supportive of greater updating for guided than unguided participants, with a larger effect occurring for younger adults. However, both younger and older adults benefitted in terms of their memory performance from the additional encoding support provided by experimenter-guided strategy implementation. Younger adults were simply better able to make use of this guidance to also form an association between encoding strategies and test outcomes and use that information to guide their strategic confidence and effectiveness ratings (as evidenced by the larger discrepancy in their strategic ratings).

As a last point, despite a clear disconnect between explicit strategic knowledge and actual strategic implementation in the current task, older adults were able to maintain similar levels of memory performance to those of younger adults. Correlations presented in the Results indicate a perseverance of behavior in older adults with respect to their strategic selections; specifically, they appear to select a strategy to use most frequently in phase 1 and continue to do so in phase 2, regardless of guidance or any implicit feedback that might be available via task experience. It may be the case, unlike in some paired-associates tasks examining imagery-based and rote encoding methods, that older adults' perseverance in this task is not necessarily damaging to their memory performance. First, it was generally the case that older adults used distinctive encoding much more frequently than relational encoding when they were able to self-select strategies, and, second, when they chose to use relational encoding, they expended a great deal of time and, presumably, effort to create a relevant relational mediator, which may compensate for the normative ineffectiveness of the strategy to some degree.

4.4 Conclusions

This research is consistent with the growing body of data that shows that both younger and older adults can benefit from a combination of relational and item-specific (i.e., distinctive) encoding (Smith, 2006; 2011). More importantly, this study offers some of the first direct evidence that younger adults can perceive the normative differential effectiveness of each strategy during practice, whereas older adults experience more difficulty translating task performance into accurate estimates of strategic effectiveness. Both age groups were better able to differentiate their strategic ratings following explicit strategic guidance in an initial study-test phase, though the effects on strategic confidence and effectiveness were larger for younger adults. Younger adults also typically entered the task with knowledge that distinctive encoding normally leads to superior memory outcomes than purely relational encoding, whereas older adults initially believed that both strategies were similarly effective. The result of task experience--especially that which was initially experimenter-guided--was to allow older adults to perceive a small degree of differential strategic effectiveness, whereas younger adults were shown that their initial impressions were correct, but the actual difference in effectiveness was larger than they previously believed. Without guidance, older adults demonstrated no ability to learn about differential strategic effectiveness, though in the current work this was not indicative of a previously-hypothesized inferential deficit, but, rather, was a consequence of memory performance actually being similar between the two strategies. An examination of the absolute accuracy of confidence postdictions revealed that both younger and older adults monitored their memory performance with a similar degree of accuracy, though their beliefs did not align with their actual memory performance. Instead, they were approximately 40 percent overconfident in their memory ability regardless of study strategy or task phase. In this sense, evidence was not found for an age-related inferential deficit related to the translation of item-level memory performance into global strategy-specific confidence estimates; everyone appeared to be similarly deficient.

Future research should continue to investigate the precise mechanisms that underlie not only memory performance differences between younger and older adults in tasks that rely on the encoding and retrieval of distinctive information, but also, and perhaps more importantly, differences in strategic knowledge and implementation. If it is the case that older adults can make use of distinctive information only when tasks support its use (e.g., Smith, 2011), it is of great importance to determine which specific components of a task are actually facilitators. Furthermore, it is necessary to investigate the apparent inferential deficit on the part of older adults: why is it that they appear to be less capable of translating their actual performance into metacognitive information that could be used to improve subsequent performance through the metacognitive self-regulation of study and retrieval behavior? For example, it would be worthwhile to explore if explicit item-level memory performance feedback associating study strategy with recollection accuracy would help older adults overcome this deficit. Behavioral (i.e., strategic) deficits will only enhance observed age-related differences in episodic memory performance, so it is important to determine if (and how) they may be overcome.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation: II*, 249. doi: 10.1016/S0079-7421(08)60422-3

Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging, 24,* 82-92. doi: 10.1037/a0014078

Battig, W, F, Montague, W. E. (1969). *Category norms for verbal items in 56 categories: A replication and extension of the Connecticut norms using University of Maryland and Illinois students (Tech. Rep.).* Boulder, CO: University of Colorado. doi: 10.1037/h0027577

Basden, B. H., Basden, D. R., & Bartlett, K. (1993). Memory and organization in elderly subjects. *Experimental Aging Research, 19,* 29-38. doi: 10.1080/03610739308253921

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. Journal of Memory and Language, 28, 610-632. doi: 10.1016/0749-596X(89)90016-8

Bieman-Copeland, S., & Charness, N. (1994). Memory knowledge and memory monitoring in adulthood. *Psychology and Aging, 9,* 287-302. doi: 10.1037//0882-7974.9.2.287

Brainerd, C. J., & Reyna, V. F. (2012). Recollective and nonrecollective recall. *Journal of Memory and Language, 63,* 425-445. doi: 10.1016/j.jml.2010.05.002

Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in older and younger adults. *Psychology and Aging, 3,* 249-257. doi: 10.1037//0882-7974.3.3.249

Brown, G. H. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1539-1553.

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

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,* 129-159.

Carroll, M. & Nelson, T. O. (1993). Effect of overlearning on the feeling of knowing is more detectable in within-subject than between-subject designs. *The American Journal of Psychology, 106,* 227-235. doi: 10.2307/1423169

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12, 50*-71. doi: 10.1037//0882-7974.12.1.50

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.) *Human memory and cognitive capabilities* (pp. 409 - 422). Amsterdam: Elsevier.

Craik, F. I. M. (2006). Remembering items and their contexts: Effects of aging and divided attention. In H. Zimmer, A. Mecklinger, & U. Lindenberger (Eds.), *Binding in human memory: A neurological perspective* (p. 571-594). Oxford University Press.

Craik, F. I. M., & Jacoby, L. L. (1979). Elaboration and distinctiveness in episodic memory. In. L.G. Nilsson (Ed.), *Perspectives on memory research* (pp.145-166). Hillsdale, NJ: Erlbaum.

Crowley, K. U., Shrager, J., & Siegler, R. S. (1997). Strategy discovery as a competitive negotiation between metacognitive and associative mechanisms. *Developmental Review, 17,* 462-489.

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

Dobbins, I. G., Kroll, N. E. A., Yonelinas, A. P., & Liu, Q. (1998). Distinctiveness in recognition and free recall: The role of recollection in the rejection of the familiar. *Journal of Memory and Language, 38,* 381-400.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25,* 691-700. doi: 10.3758/BF03211311

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? Psychology and Aging, 13, 597-607. doi: 10.1037//0882-7974.13.4.597

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15,* 462-474. doi: 10.1037//0882-7974.15.3.462

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition, 29,* 247-253. doi: 10.3758/BF03194918

Epstein, M. L., Phillips, W. D., Johnson, S. J. (1975). Recall of related and unrelated word pairs as a function of processing level. Journal of Experimental Psychology: Human Learning and Memory, 1, 149-152. doi: 10.1037//0278-7393.1.2.149

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34,* 906-911. doi: 10.1037//0003-066X.34.10.906

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology; Learning, Memory, and Cognition, 33,* 238-244. doi: 10.1037/0278-7393.33.1.238

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58,* 19-34. doi: 10.1016/j.jml.2007.03.006

Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging can spare recollection-based retrieval monitoring: The importance of event distinctiveness. *Psychology and Aging, 22,* 209-213.

Gardiner, J. M., & Klee, H. (1976). Memory for remembered events: An assessment of output monitoring in free recall. *Journal of Verbal Learning and Verbal Behavior, 15,* 227–233. doi: 10.1016/0022-5371(76)90021-9

Geraci, L., & Rajaram, S. (2002). The orthographic distinctiveness effect on direct and indirect tests of memory: Delineating the awareness and processing requirements. *Journal of Memory and Language, 47,* 273-291. doi: 10.1016/S0749-596X(02)00008-6

Geraci, L., McDaniel, M. A., Manzano, I., Roediger, H. L., (2009). The influence of age on memory for distinctive events. *Memory & Cognition*, *37,* 175 - 180. doi: 10.3758/MC.37.2.175

Hay, J. F., Jacoby, L. L. (1999). Separating habit and recollection in young and older adults: Effects of elaborative processing and distinctiveness. *Psychology and Aging, 14,* 122-134. doi: 10.1037//0882-7974.14.1.122

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

Hertzog, C. (2008). Theoretical approaches to the study of cognitive aging: An individual differences perspective. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook of cognitive aging* (pp. 34-49). Thousand Oaks, CA: Sage Publications.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task-performance in adults. *Psychology and Aging, 5,* 215-227. doi: 10.1037//0882-7974.5.2.215

Hertzog, C., & Dunlosky, J. (2004). Aging, metacognition, and cognitive control. *The Psychology of Learning and Motivation, 45,* 215-251. doi: 10.1016/S0079-7421(03)45006-8

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and old age (pp. 417-466). *Handbook of Aging and Cognition II*. Mahwah, NJ: Erlbaum.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is used for judgments about learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29, 22-34. doi: 10.1037/0278-7393.29.1.22

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17,* 209-225. doi: 10.1037//0882-7974.17.2.209

Hertzog, C., Price, J., Burpee, A., Frentzel, W. J., Feldstein, S, & Dunlosky, J. (2009). Why do people show minimal knowledge updating with task experience: Inferential deficit or experimental artifact? *The Quarterly Journal of Experimental Psychology, 62,* 155-173. doi: 10.1080/17470210701855520

Hertzog, C., Price, J., & Dunlosky, J. (2008). How is knowledge generated about memory encoding strategy effectiveness? *Learning and Individual Differences, 18,* 430-445. doi: 10.1016/j.lindif.2007.12.002

Hertzog, C., Price, J., & Dunlosky, J. (2012). Age differences in the effects of experimenter-instructed versus self-generated strategy use. *Experimental Aging Research, 38,* 42-62. doi: 10.1080/0361073X.2012.637005

Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory & Cognition, 30,* 67-80. doi: 10.3758/BF03195266

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging, 24,* 462-475. doi: 10.1037/a0014417

Hirshman, E., Whelley, M. M., & Palij, M. (1989). An investigation of paradoxical memory effects. *Journal of Memory and Language*, *28*, 594 - 609. doi: 10.1016/0749-596X(89)90015-6

Hultsch, D. F. (1974). Learning to learn in adulthood. *Journal of Gerontology, 29,* 302-308. doi: 10.1159/000271531

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review, 2,* 105-112. doi: 10.3758/BF03214414

Hunt, R. R. (2002). Two contributions of distinctive processing to accurate memory. *Journal of Memory and Language, 48,* 811-825.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32,* 421-455. doi: 10.1006/jmla.1993.1023

Hunt, R. R., & Seta, C. E. (2011). Knowledge affords distinctive processing in memory. *Journal of Memory and Language, 65,* 390-405. doi: 10.1016/j.jml.2011.07.001

Hunt, R. R., & Seta, C. E. (1984). Category size effects in recall: The roles of relational and individual item information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 454-464. doi: 10.1037//0278-7393.10.3.454

Jackson, D. K., & Schneider, H. G. (1982). Age differences in organization and recall: An analysis of rehearsal processes. *Psychological Reports, 50,* 919-924. doi: 10.2466/pr0.1982.50.3.919

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language 30,* 513–541. doi: 10.1016/0749-596X(91)90025-F

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, Neurospsychology, & Cognition, 12,* 278-298. 10.1080/138255890968312

Hunt, R. R., & Smith, R. E. (1996). Accessing the particular from the general: The power of distinctiveness in the context of organization. *Memory & Cognition, 24,* 217-225. doi: 10.3758/BF03200882

Kausler, D. H. (1994). *Learning and memory in normal aging*. San Diego, CA, US: Academic Press.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349-370. doi: 10.1037//0096-3445.126.4.349

Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learner's sensitivity to retrieval conditions at test. *Memory & Cognition, 34,* 959-972.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. The Journal of Experimental Psychology: General, 123, 297-315.

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

Koriat, A., Goldsmith, M., & Halamish, V. (2008). Controlled processes in voluntary remembering. In H. L. Roediger, III (Ed.), Cognitive psychology of memory. Vol. 2 of Learning and memory: A comprehensive reference, 4 Vols. (J. Byrne, Editor) (pp. 307-324). Oxford: Elsevier.

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

Koutstaal, W. (2006). Flexible remembering. *Psychonomic Bulletin & Review, 13,* 84-91. doi: 10.3758/BF03193817

Laurence, M. W. (1966). Age differences in performance and subjective organization in the free-recall learning of pictorial material. Canadian Journal of Psychology, 20, 388-399. doi: 10.1037/h0082958

Littell, R. C, Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (2000). *SAS system for mixed models (4th. Ed.)*. Cary, NC: SAS Institute, Inc.

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 756-766. doi: 10.1037/0278-7393.10.4.756

Lovelace, E. A. (1990). Aging and cognition: Mental processes, self-awareness, and interventions. *Advances in Psychology, 72,* 407-434. doi: 10.1016/S0166-4115(08)60782-7

Luo, L., & Craik, F. I. M. (2008). Aging and memory: A cognitive approach. *The Canadian Journal of Psychiatry, 53,* 346-353.

Luo, L., & Craik, F. I. M. (2009). Age differences in recollection: Specificity effects at retrieval. *Journal of Memory and Language, 60,* 421-436. doi: 10.1016/j.jml.2009.01.005

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,* 242-249. doi: 10.1037//0882-7974.5.2.242

Manis, M., Shedler, J., Jonides, J., & Nelson, T. E. (1993). Availability heuristic in judgments of set size and frequency of occurrence. *Journal of Personality and Social Psychology, 65,* 448-457.

Mäntylä, T. (1986). Optimizing cue effectiveness: Recall of 500 and 600 incidentally learned words. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14,* 66-71. doi: 10.1037//0278-7393.12.1.66

Mäntylä, T., & Nilsson, L. (1988). Cue distinctiveness and forgetting: Effectiveness of self-generated retrieval cues in delayed recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 502-509. doi: 10.1037//0278-7393.14.3.502

Matvey, G., Dunlosky, J., Shaw, R. J., Parks, C., & Hertzog, C. (2002). Age-related equivalence and deficit in knowledge updating of cue effectiveness. *Psychology and Aging, 17,* 589-597. doi: 10.1037//0882-7974.17.4.589

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

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

McDaniel, M. A., Dornburg, C. C., & Guynn, M. J. (2005). Disentangling encoding versus retrieval explanations of the bizarreness effect: Implications for distinctiveness. *Memory & Cognition, 33,* 270-279. doi: 10.3758/BF03195316

McDaniel, M. A., Einstein, G. O., DeLosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology, Learning. Memory, and Cognition, 21,* 422-435. doi: 10.1037//0278-7393.21.2.422

McDaniel, M. A., Einstein, G. O., & Jacoby, L. L. (2008). New considerations in aging and memory: The glass may be half full. In F. I. M. Craik & T. A. Salthouse (Eds.) The handbook of aging and cognition (3rd ed., pp. 251-310). New York: Psychology Press.

McDaniel, M. A., & Geraci, L. (2006). Encoding and retrieval processes in distinctiveness effects: Toward an integrative framework. In R. R. Hunt & J. B. Worthen (Eds.) *Distinctiveness and Memory* (pp. 65-88). Oxford University Press. doi: 10.1093/acprof:oso/9780195169669.003.0004

Medin, D. L., Goldstone, R. L., & Genter D. (1993). Respects for similarity. *Psychological Review, 100,* 254-278. doi: 10.1037//0033-295X.100.2.254

Murphy, M. D., Schmitt, F. A., Caruso, M. J., & Sanders, R. E. (1987). Metamemory in older adults: The role of monitoring in serial recall. *Psychology and Aging, 2,* 331-339. doi: 10.1037//0882-7974.2.4.331

Murphy, K. J., West, R., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2007). Word-list-learning preformance in younger and older adults: Intra-individual performance variability and false memory. *Aging, Neuropsychology, and Cognition, 14,* 70-94. doi: 10.1080/138255890969726

Naveh-Benjamin, M., Craik, F. I. M. (1995). Memory for context and its use in item memory: Comparisons of younger and older persons. *Psychology and Aging, 10,* 284-293.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 84*, 93-116. doi: 10.1037//0033-2909.95.1.109

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (E.) *The psychology of learning and motivation* (pp. 125-173). New York: Academic Press.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning are extremely accurate at predicting subsequent recall: The delayed JOL effect. *Psychological Science, 2,* 267-270.

Price, J., Hertzog, C., & Dunlosky, J. (2008). Age-related differences in strategy knowledge updating: blocked testing produces greater improvements in metacognitive accuracy for younger than older adults. *Aging, Neuropsychology, and Cognition, 15(5), 601-626.* doi: 10.1080/13825580801956225

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

Rabinowitz, J. C., Freeman, K., & Cohen, S. (1992). Use and maintenance of strategies: The influence of accessibility to knowledge. *Journal of Educational Psychology, 84,* 211-218.

Rankin, J. L., & Collins, M. (1986). The effects of memory elaboration on adult age-differences in incidental recall. *Experimental Aging Research, 12,* 231-234. doi: 10.1080/03610738608258574

Robinson, A. E., Hertzog, C., & Dunlosky, J. (2006). Aging, encoding fluency, and metacognitive monitoring. *Aging, Neuropsychology, and Cognition, 13,* 458-478.

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

Sahakyan, L., Delaney, P. F., & Goodman, L. B. (2008). Oh, honey, I already forgot that: Strategic control of directed forgetting in older and younger adults. *Psychology and Aging, 23,* 621-633.

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

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19,* 523-542. doi: 10.3758/BF03197149

Schacter, D. L., Buckner, R. L., Koustaal, W., Dale, A. M., & Rosen, B. R. (1997). Late onset of anterior prefrontal activity during true and false recognition: And event-related fMRI study. *Neuroimage, 6,* 259-269.

Schacter, D. L., Reiman, E., Curran, T., Sheng Yun, L., Bandy, D., McDermott, K. B., & Roediger, H. L. III (1996). Neuroanatomical correlates of verdical and illusory recognition memory: Evidence from positron emission tomography. *Neuron, 17,* 267-274.

Schunn, C. D., & Reder, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General, 130,* 59-76.

Sego, S. A., Golding, J. M., & Gottlob, L. R. (2005). Directed forgetting in older adults using the items and list methods. *Aging, Neuropsychology, and Cognition, 13,* 1-20. doi: 10.1080/138255890968682

Singer, J. D. (1998). Using SAS proc mixed to fit multi-level models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24,* 323-355.

Smith, R. E. (2006). Adult age differences in memory: Item-specific, organizational, and distinctive information. In R.R. Hunt & J.B. Worthen (Eds.) *Distinctiveness and memory* (pp. 259 - 287). New York: Oxford University Press.

Smith, R. E. (2011). Providing support for distinctive processing: The isolation effect in younger and older adults. *Psychology and Aging, 26,* 744-751. doi: 10.1037/a0022715

Smith, R. E., & Engle, R. W. (2011). Study modality and false recall: The influence of resource availability. *Experimental Psychology, 58,* 117-124. doi: 10.1027/1618-3169/a000076

Smith, R. E., & Hunt, R. R. (2000). The effects of distinctiveness require reinstatement of organization: The importance of intentional memory instructions. *Journal of Memory and Language, 43,* 431-446. doi: 10.1006/jmla.2000.2707

Stine-Morrow, E. A. L., Gagne, D. D., Morrow, D. G., & DeWall, B. H. (2004). Age differences in rereading. *Memory & Cognition, 32,* 696-710. doi: 10.3758/BF03195860

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6,* 662-667. doi: 10.3758/BF03212976

Tullis, J. G., & Benjamin, A. S. (2012). The effectiveness of updating metacognitive knowledge in the elderly: Evidence from metamnemonic judgments of word frequency. Psychology and Aging, 27, 683-690. doi: 10.1037/a0025838

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2012). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition.* Advance online publication. doi: 10.3758/s13421-012-0274-5

Van Overschelde, J. P., Rawson, K. A., and Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289-335. doi: 10.1016/j.jml.2003.10.003

Verhaeghen, P., Marcoen, A., & Goossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. *Journals of Gerontology: Psychological Sciences, 48,* 157-171.

Von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologishe Forschung, 18,* 299-342. doi: 10.1515/bchm2.1912.77.5.303

Waddill, P. J., McDaniel, M. A. (1998). Distinctiveness effects in recall: Differential processing or privileged retrieval? *Memory & Cognition, 26,* 108-120. doi: 10.3758/BF03211374

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale – Revised.* New York: Psychological Corporation.

White, K. G., & Wixted, J. T. (1999). Psychophysics of remembering. *Journal of Experimental Analysis of Behavior, 71,* 91-113.

Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences, 8,* 327-353.

Witte, K. L., Freund, J. S., & Sebby, R. A. (1990). Age differences in free recall and subjective organization. *Psychology and Aging, 5,* 307-309. doi: 10.1037//0882-7974.5.2.307

Zachary, R. (1986). *Shipley Institute of Living Scale – Revised manual.* Los Angeles: Western Psychological Services.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.) *The handbook of cognition and aging* (2nd ed. pp. 293 - 357). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.