

Running Head: AGE AND REDINTEGRATION

Age and Redintegration in Immediate Memory and Their Relationship to Task Difficulty.

Kerry Neale

Gerald Tehan

University of Southern Queensland, Toowoomba, Australia

Address correspondence to

Gerry Tehan

Department of Psychology

University of Southern Queensland

PO Box 4196, Springfield, QLD, 4300

Australia

e-mail: [tehan@usq.edu.au](mailto:tehan@usq.edu.au)

Fax: +61 7 46312721

### Abstract

It is commonly assumed that as short-term memory tasks become more difficult a transient phonological trace that supports recall loses its fidelity. Recall can still be achieved through a process called redintegration where long-term phonological or lexical knowledge is used to reconstruct the memory trace. In the current research we explore age-related differences on the redintegration process by having older and younger participants study lists under different levels of task difficulty. In experiment 1 semantic similarity was manipulated and in Experiment 2 phonological similarity was varied as a means of examining the redintegration process. The results show that similarity effects can be accurately predicted from knowledge of task difficulty, with item scoring but not for order scoring. The results indicate support for the redintegration perspective and that while there may be differences in the absolute level of recall across age groups the redintegration process is identical for younger and older participants.

There is general agreement in the literature that normal aging brings measurable declines in cognitive performance. Age-related decrements are readily evident in long-term episodic recall tasks, but are also prevalent in short-term retention tasks as well. For instance, robust aging decrements are found in complex working memory tasks (e.g. reading span, operation span, counting span) where performance is seen to be determined by the joint function of storage and processing requirements. Explanations for such results have tended to concentrate on the processing aspect and have thus been couched in terms of differences in cognitive resources ( Craik & Byrd, 1982), cognitive slowing (Salthouse, 1996), and in problems with executive functioning such as failures to inhibit irrelevant material (Hasher & Zacks, 1988; Rouleau & Belleville, 1996), problems in coordinating tasks and information streams (Kramer, Hahn, & Gopher, 1999), and in costs associated with task switching, (Mayr, Spieler, & Kliegl, 2001). Age-related differences in storage requirements in these tasks have been largely ignored, which is somewhat surprising given the widespread belief that complex span and simple span tasks share a common storage component (Cowan, 1999; Engle, Kane, & Tuholski, 1999; Colom, Shih, Flores-Mendoza, & Quiroga, 2006) and that reliable age differences in simple span tasks do exist (Bopp & Verhaeghan, 2005). One exception to this trend is the work of Oberauer and his colleagues (Oberauer, 2001, 2005a, 2005b; Oberauer & Kliegl, 2001) who have explored age and working memory issues within Cowan's (1995) "focus of attention" model of short-term storage. While Oberauer has concentrated upon a specific model, the aim of the current research is to focus upon general principles rather than any specific model. Our interest is in one widely held assumption that is common to many models of short-term/working memory. That is, long-term memory is used to reconstruct degraded short-term traces; a process known as redintegration.

Most current models of immediate serial recall (simple span) posit a two part process to recall. The first step is for some form of order retrieval mechanism to produce a phonological

representation of a candidate for output. It is widely assumed that at study speech-based representations of items are established and that in the absence of rehearsal those representations lose their fidelity either through decay (Baddeley, 1986; Burgess & Hitch, 1996; Page & Norris, 1998; Henson, 1998) or through interference (Nairne, 1990; Tehan & Humphreys, 1995). At recall it is assumed that the memory trace may or may not be degraded. If the trace is intact then recall will not be problematic. However, if the trace is degraded a second step is initiated. Long-term lexical/phonological information is accessed in the hope that such information can be used to reconstruct the item (e.g. using knowledge about words to generate a word from a fragment like *cr\_ \_odi\_e*). This reconstruction processes is often referred to as redintegration (Brown & Hulme, 1995; Schweickert, 1993).

Schweickert (1993) formally tested these ideas by developing a multinomial processing tree model of immediate recall. The model assumes that three outcomes are possible when attempting to recall an item. The first assumes that there is a certain probability,  $I$ , that representation of an item is intact and a correct response will be produced. The second assumes that there is a certain probability,  $R$ , that memory trace is degraded but a correct response can still be generated through the redintegration process. Finally, there is a certain probability that the trace is degraded and that redintegration is unsuccessful resulting in an error. With these assumptions the probability of correctly recalling an item is given by the equation  $I + (1-I)R$ . That is, it is the sum of the probability of retrieving the item intact, plus the product of the probability of item being not intact with the probability that it can be reconstructed.

Schweickert, Chen, and Poirier (1999) explored these ideas by manipulating factors that were assumed to influence the degree to which the trace would remain intact and those factors that were thought to influence redintegration. For example, they argued that serial position and word length would impact upon the integrity of the trace, but long-term lexical

factors like word frequency and lexicality would determine how easily a degraded representation could be reconstructed. The computational model provided very good fits of relevant empirical data.

The research reported here explores age differences in the recall process by using the Schweickert et al. (1999) logic. That is, we intend to manipulate factors that are assumed to affect the fidelity of the memory trace and factors that are assumed to influence redintegration. To this end, the empirical literature on immediate serial recall has consistently demonstrated that recall is better if items are read aloud than if read silently, if list length is short rather than long, and if memory is tested immediately rather than after a filled retention interval. Within most current models of immediate memory, auditory modality either ensures registration in a short-term store (Baddeley, 1986) or results in stronger (Tolan & Tehan, 1999) or more discriminative representations (Nairne, 1990). List length is often used as a mechanism for utilising short-term storage capacity. From a rehearsal/decay perspective, the greater the number of items in a list, the less chance there is of any decaying representation being refreshed by rehearsal. Employing a retention interval that is filled with rehearsal preventing distractor activity has likewise been a traditional means of ensuring that the memory trace is degraded either through decay or via retroactive interference. In short, all of these factors are assumed to influence the degree of degradation of the memory trace. Thus, it is argued that the trace for an item that has been read aloud in a four item list that is tested immediately is more likely to be intact than a visually presented item in a six word list that is tested after a four-second filled delay. This latter representation is likely to benefit from redintegration if redintegration is possible.

Our measure of redintegration involves similarity among the to-be-remembered items and is based upon suggestions first proposed by Poirier and Saint-Aubin (1995; Saint-Aubin & Poirier, 1999a; 1999b). Poirier and Saint-Aubin (1995) re-examined the widely held idea

that similarity amongst list items in immediate serial recall had an adverse effect upon order memory. While this finding is relatively consistent when phonological similarity is manipulated, Poirer and Saint-Aubin argued that this was not necessarily the case with semantic similarity. In their experiments that explored semantic similarity effects on order memory, each semantically similar study list contained items from the one taxonomic category; items in a dissimilar list came from different taxonomic categories. When the task involved immediate serial recall and performance was scored as correct only if the item was recalled in its correct serial position, they found that recall for the semantically similar lists was better, not worse, than for the dissimilar lists. To understand these results they then made a more detailed examination of their data by using separate item and order scoring techniques. In item scoring an item is scored as correct if it is output irrespective of what serial order it is recalled in. An order error occurs when an item is recalled, but recalled in the wrong serial position. In order scoring, these order errors are conditionalised upon item scores (Saint-Aubin & Poirier, 1999b) or order accuracy measures are derived by conditionalising correct-in-position scores on item scores (Fallon, Groves, & Tehan, 1999). Using these alternate scoring procedures, the similarity advantage was evident using item scores, but there was no difference between similar and dissimilar scores for order memory.

Poirier and Saint-Aubin (1995) explained their results in what amounts to a redintegration argument. Like others they assumed that at study a phonological trace of the items was created and through decay or interference this trace became degraded at recall. Item errors were attributed to failure of the reconstruction process, but such failures were seen to be dependent upon firstly, the degree of degradation in the phonological trace and secondly, the availability of a long-term memory representation. Thus, given a degraded trace, participants would search their long-term memories for a potential candidate for recall. With lists of items from the same semantic category, category knowledge could be used to

restrict the size of the search area in long-term memory leading to an increased likelihood that an item would be recovered (e.g. knowing that all the items on the list were reptiles could facilitate the reconstruction of a fragment like *cr\_\_odi\_e*). Order errors were attributed to problems in discrimination, not in problems of retrieval or reconstruction. That is, if item representations were similar, a degraded phonological trace could be matched to a number of potential candidates, once such candidates had been generated. The fact that, unlike phonological similarity, semantic similarity had no detrimental effect upon order memory was explained by arguing that the phonological representations of semantically similar and dissimilar items were equally discriminable. In short, redintegration effects, according to this account, are reflected in item scoring and are due to similarity being used as a cue to facilitate item recovery.

Applying the Poirier and Saint-Aubin ideas allows us to explore age differences in the redintegration process. Kausler (1994) reviewed much of the literature on age differences in short-term memory performance. That review established that while there are quantitative age differences in modality, capacity and distractor activity, there are no qualitative differences. Patterns of modality and suffix effects were equivalent for younger and older participants, exceeding span had similar effects and forgetting rates were the same for both age groups. This suggests that there are no fundamental changes in short-term memory processing across age groups and as a consequence one might predict that the determinants of task difficulty would be age invariant. This does not mean that there might still remain relative differences across age groups. Thus both age groups might find recall of a six item list harder than a four item list but older participants may still perform more poorly on both lists. Thus, as a working hypothesis, we assume that aging, for some unspecified reason, results in a reduced likelihood that a short-term memory trace will be intact at the point of retrieval and redintegration processes would be required to facilitate recall. However, once redintegration

was required, those same processes would be involved for all age groups. That is, the redintegration process is likely to be age invariant as well. In the case where semantic similarity is involved, reducing the search set in memory via a category cue would facilitate recall for younger and older participants alike.

### Experiment 1

In the current experiment younger and older participants studied lists of semantically similar or dissimilar items and serial recall of these items was requested. These lists were visually presented on a computer screen and were either read aloud or read silently; the lists were either four items in length or six items in length and memory for each list was tested either immediately, after two seconds of digit shadowing or after four seconds of digit shadowing. Robust modality, list length and retention interval effects should be readily apparent. Age and similarity effects are the primary variables of interest.

Given prior research it is expected that age-related differences and semantic similarity effects will emerge. However, rather than concentrate upon mean differences, our interest is focused upon redintegration effects. The expectation is that as task difficulty increases redintegration effects will become stronger. These expectations, however, immediately present two problems. The first is that there is the tacit implication that task difficulty exists upon a continuum. The second is that there is no common means of operationalising task difficulty when multiple manipulations of difficulty are attempted.

In its simplest form of the redintegration model appears to assume that task difficulty can be expressed on a single dimension from easy to difficult. Schweickert et al., (1999) in a number of computational tests of the multinomial model used serial position as their manipulation of task difficulty. Standard serial position curves (with the possible exception of the terminal item) conform to the continuum assumption, with the first item being recalled better than the second and the second better than the third and so on. In another instance they



manipulated word length as their measure of task difficulty. Again the assumption that one syllable, two syllable, and three syllable words lie upon a continuum of difficulty seems reasonable.

The above examples all involve a single variable associated with task difficulty. The problem surfaces when multiple methods of difficulty are simultaneously manipulated. As we have indicated previously, the choice of modality, list length, and retention interval were selected not only because they are known to impact upon serial recall performance, but also because they are assumed to influence different memory processes. According to some models modality influences registration in memory whereas list length and retention intervals affect storage and forgetting factors. Thus, it is clearly implicit that task difficulty is multi-faceted but it is also reasonable to expect that combining factors is likely to produce more disruption than that achieved by presenting each factor alone. Consequently, even though task difficulty may be multiply determined, it still may be possible to derive an index of task difficulty that lies on a continuum. For instance, the top panel of Figure 1 presents the results of a hypothetical 2x2x3 experiment in the traditional manner. Main effects and interactions are easily derived but it is clear that some conditions are harder than others. In the bottom panel of Figure 1, the same data have been presented but in a way that is consistent with the redintegration assumptions. That is, the results are presented in rank order of higher levels of recall to lower levels of recall. The problem here though is how does one determine the ranking of the various conditions?

-----  
Insert Figure 1 about here  
-----

There appear to be three possible ways of ranking the conditions: Take the average of the similar and dissimilar lists, use the dissimilar lists as baseline, or use the similar lists as baseline. Provided that modality and retention interval effects are equivalent for similar and

dissimilar lists, any of the three measures should be adequate. In any event, rank order correlations can be used to confirm or disconfirm that task difficulty is equivalent across similarity conditions.

In addition, instead of presenting mean performance for the two similarity conditions, the outcomes of the redintegration process can be depicted in terms of the size of the similarity effect; the difference between similar and dissimilar conditions. The triangles on Figure 1 represent the magnitude of the similarity effect in each condition (as noted on the right hand Y axis). Note that as task performance appears to deteriorate in this hypothetical data, the magnitude of the similarity effect increases. Plotting a best-fit trendline seems to be an appropriate way to evaluate the redintegration hypotheses of our study.

In the current experiment the 2x2x3 (modality, list length, retention interval) design produces 12 estimates of task difficulty. While any of the three means of operationalising task difficulty could have been used, performance on the dissimilar lists is being used as the baseline. The assumption is that with dissimilar lists, there is nothing available to the participant that will facilitate redintegration. That is, participants will have to access all of their long-term lexical/phonological memory to reconstruct a degraded item. It seems to us that this is the logical baseline from which to assess the hypothesised advantages of the redintegration process. Consequently, we are using the number of errors on the dissimilar lists as our measure of task difficulty. The zero point on the x-axis represents errorless performance. A score of .2 on the task difficulty dimension would mean that, on average, 20% of the recalls on the dissimilar lists were errors of one form or another. Thus, we are using performance on the dissimilar lists as a baseline measure and at each level of task difficulty we are measuring the size of the similarity advantage. A redintegration effect would be represented as an increase in the similarity advantage as errors on the dissimilar lists increase.

## Method

### *Participants*

Forty volunteers, either from the community or psychology undergraduates from the University of Southern Queensland, participated for tickets in the departmental prize draw or course credit. The sample of convenience consisted of 20 younger adults with an age range of 18 to 39 years ( $M = 22.05$ ,  $SD = 6.10$ ) and twenty older adults with an age range of 61 to 79 years ( $M = 68.90$ ,  $SD = 5.23$ ). All participants lived independently in the community, were native Australian English speakers, and had normal or correct-to-normal vision and hearing. Both participant groups reported themselves to be in good health and to have no difficulty reading the words as they would be presented on the computer screen. The mean number of years of education did not differ between younger adults ( $M = 12.95$ ,  $SD = 2.31$ ) and older adults ( $M = 11.20$ ,  $SD = 3.22$ ),  $F(1,38) = 3.90$ . All participants were administered the Wechsler Test of Adult Reading (WTAR; Wechsler, 2001) in order to obtain an estimate of each participants overall level of cognitive intellectual functioning. The WTAR is a brief reading test that is highly correlated with WAIS-III IQ. There was no difference between younger (Mean Estimated IQ = 108.95,  $SD = 9.55$ ) and older adults (Mean Estimated IQ = 111.20,  $SD = 10.67$ ) in their estimated WAIS-III full scale IQ,  $F(1,38) = 0.49$ .

### *Materials*

Participants studied four blocks of thirty trials in two one-hour sessions, the sessions being separated by a week. These blocks were presented in a fixed order. The first block consisted of four-item lists that were read silently. Following a brief break the second block of six-item lists read silently was then presented. A week later, the third block of four-item lists read aloud was studied followed by the fourth block of six-item lists read aloud. Each block consisted of 30 trials, 15 semantically similar (e.g., *chicken, horse, goat, sheep, duck, hen*) and 15 semantically dissimilar (e.g., *shoe, glass, fan, tree, car, map*). The 15 trials

include 5 trials for immediate recall, 5 trials for recall after a 2-s filled delay, and 5 trials for recall after a 4-s filled delay.

The same words were used in each of the four blocks but were randomly assigned to the different conditions. The 180 words used as experimental stimuli consisted of six medium strength instances from each of 30 different medium sized taxonomic categories, selected from the University of South Florida taxonomic category norms (McEvoy & Nelson, 1982). The choice of medium sized categories with medium strength instances was made to minimise the chance that participants could simply guess the answer, but at the same time making sure that the instances were well known members of the category.

To create the 30 trials in each of block-1 and block-2, the categories were first randomised and then the six instances within each category were also randomised. Instances from the first 15 categories were simultaneously allocated to the semantically similar trials in the four-word condition (block-1), and to the dissimilar trials in the six-word lists (block-2). Likewise, instances from the last 15 categories were allocated to the dissimilar trials on the four-word lists (block-1) and the similar trials in the six-word lists (block-2). To construct the dissimilar trials, the ninety words were randomised and then randomly assigned to the 15 trials. In the four-word lists, the first four of the six items were selected to be the list items. Each item was therefore sampled twice and appeared in an intact category in the four-word condition and in a dissimilar category in the six-word condition, or appeared in a dissimilar category in the four-word condition and in an intact category in the six-word condition. Randomly generated numbers between 10 and 99 were allocated to trials selected to be recalled after a delay. Two 2-digit numbers were allocated to each of the 2-s delay lists and four 2-digit numbers were allocated to the 4-s delay lists. The order of the 30 trials in both the four-word and six-word condition was then randomised.

The above procedure was repeated to generate new trials for Blocks 3 and 4.

### *Procedure*

Participants completed three practice trials prior to presentation of each experimental block. The practice trials were presented in the same manner as the experimental trials. When the experimenter was satisfied that participants could perform each task, testing commenced.

Participants were tested individually. Each trial began with an audible beep, followed one second later with a second beep in conjunction with presentation of the word “READY” in uppercase. The experimental stimuli were then presented in lower case in the centre of a computer screen at the rate of one word per second. These items were presented visually and were either read silently by the participant or read aloud depending upon the experimental condition. In all conditions, if numbers appeared on the screen they too were presented at a rate of one digit pair per second and participants were instructed to say the digit pair aloud as they appeared on the screen (e.g. “*sixty-four, twenty-two*”). At the end of each trial, a row of question marks (????) appeared as a prompt at which point the participant attempted to verbally recall the items in the order in which they were presented. To ensure that output order was maintained, participants were advised to substitute the word “pass” for any word in the list that could not be recalled. The next trial commenced after a 12-second delay during which participants attempted to recall the presented items. The experimenter recorded the responses on a hard copy of the input file.

## Results

### *Scoring*

Serial recall is traditionally scored by considering as correct only those items that have been recalled in the same serial position as they were presented in. From this perspective omissions, order errors and intrusions of any type constitute an error. However, in recent times two alternative procedures have been employed. Item scoring ignores the serial position information and scores as correct any item from the list that has been recalled. From this

perspective omissions and intrusions are the only types of errors. Order accuracy is measured by simply dividing the correct-in-position score by the item score. This score reflects the proportion of items that were correctly recalled in position, given that the item was recalled in the first place. All three scoring procedures are reported here.

#### *Age effects in mean levels of performance*

The experimental design was a 2 (age) x 2 (similarity) x 2 (list length) x 2 (modality) x 3 (retention interval) mixed design, with age being the sole between-subjects variable. The table of means are presented in Appendix A. For current purposes, the main finding of interest is that there was a significant main effect for age for correct-in-position,  $F(1,38) = 13.99$ ,  $MSe = .17$ ,  $p. < .001$ ; item scoring,  $F(1,38) = 8.93$ ,  $MSe = .13$ ,  $p. < .01$ ; and order accuracy scoring,  $F(1,38) = 10.68$ ,  $MSe = .16$ ,  $p. < .001$ . In all instances, the younger participants were more accurate than older adults. Robust benchmark effects of list length, modality, and retention interval effects were also readily apparent. (The outcomes of a 2x2x2x2x3 ANOVA can be retrieved from [www.usq.edu.au/users/tehan/ageanova.doc](http://www.usq.edu.au/users/tehan/ageanova.doc).)

#### *Task Difficulty*

Our premises presuppose that there is a valid measure of task difficulty. In Table 1 we present the rank order correlations (in bold font in top left corner) among the 12 estimates of task difficulty. It is very clear that the ranking of our 12 estimates of task difficulty are very similar for similar and dissimilar lists, for younger and older participants for both correct-in-position and item scoring. The rank order for the tasks is as follows where the first digit represents list length, the letter represents read Aloud or read Silent, and the second digit represents retention interval: 4-A-0, 4-S-0, 4-A-2, 4-A-4, 6-A-0, 4-S-2, 6-A-2, 6-S-0, 4-S-4, 6-A-4, 6-S-2, and 6-S-4. For order scoring there is more variability in the measures of task difficulty. At least for the former two measures of scoring the data, we are confident that we have achieved a valid measure of task difficulty.

-----  
Insert Table 1 about here  
-----

### *Redintegration Effects*

Figure 2 depicts the size of the similarity advantage as a function of task difficulty. A positive similarity effect reflects superior recall of the similar lists relative to the dissimilar lists.

-----  
Insert Figure 2 about here  
-----

### *Correct-in-position*

As can be seen in the upper panel of Figure 2, there is a tendency for the size of the similarity effect to increase as task difficulty increases, but the strength of the relationship between task difficulty and the size of the similarity effect is quite low ( $r^2 = .55$  and  $.21$  for younger and older samples respectively). There was no difference in the slopes for younger ( $b = .15$ ) and older participants ( $b = .09$ ),  $t(20) = .97$ ,  $p. > .05$ , nor for the intercepts ( $c = .03$  and  $.05$  for younger and older participants respectively),  $t(20) = .45$ ,  $p. > .05$ . The similarity of the slopes indicates that the underlying processes are much the same for younger and older adults.

### *Item scoring*

The middle panel of Figure 2 depicts the redintegration effects for item scoring. For both younger and older groups there is a very strong relationship between task difficulty and the size of the similarity effect ( $r^2 = .88$  for younger adults and  $r^2 = .87$  for older adults). That is, as task difficulty increases the size of the similarity advantage increases in a highly predictable way. The equivalence of the slopes for older ( $b = .36$ ) and younger participants ( $b = .41$ ),  $t(20) = .71$ ,  $p. > .05$ , and intercepts ( $c = .02$  for both groups),  $t(20) = .23$ ,  $p. > .05$ ,

suggests that there is no real difference between younger and older adults in the use of redintegration.

### *Order accuracy*

Order accuracy is the proportion of items that were recalled in their correct serial position given that they were recalled somewhere on the output protocol. As can be seen in the lower panel of Figure 2, the relationship between task difficulty and the similarity effect for order information is extremely weak ( $r^2 = .10$  and  $.13$  for younger and older groups respectively). The regression line appears to be flat across all levels of task difficulty and at the zero mark. The slope and intercepts are equivalent for younger ( $b = -.09$  and  $c = .02$ ) and older participants ( $b = -.09$  and  $c = .02$ ),  $t(20) = .52, p. > .05$  and  $t(20) = .01, p. > .05$ .

### Discussion

The results of the current experiment confirm previous findings that age differences can be observed in simple short-term memory tasks. We also replicate previous findings that semantically similar lists are better recalled in position than dissimilar lists. The upper panel in Figure 2 suggests that on an immediate test, the effects of similarity are minimal, but that the effects become stronger as task difficulty increases. One ready explanation for this is that with an immediate test, the phonological trace is sufficiently intact that direct retrieval is possible or that there are no problems in discrimination.

The middle panel depicts the primary finding of the experiment. The pattern that is found with correct-in-position scoring is again apparent, but is much stronger with item scoring. It is clear, at least at the level of group means, that given the likelihood of making an error on the dissimilar lists, it is possible to predict the size of the similarity advantage at that point with some precision. The strong linear function is consistent with a redintegration perspective that asserts as task difficulty increases, there is an increased likelihood that long-term memory will be accessed and that similarity can act as a cue to narrow the number of



potential candidates for recall. Importantly, redintegration effects appear to be equivalent for younger and older participants.

The robust item effects are not replicated in the order accuracy measure where order accuracy is equivalent for both similar and dissimilar lists and centres around the zero percent advantage. As mentioned earlier, most models of memory assume that similarity involves similarity of representations. As such similarity reduces trace discriminability and thereby impacts predominantly upon memory for order. The zero percent finding suggests that the representations underpinning performance are equally discriminable for items that all come from the same taxonomic category or come from diverse categories. That is, in spite of using a label of semantic similarity, the representations of semantically similar items are no more similar to each other than items in dissimilar lists. In addition, it seems that there is no relationship between task difficulty and any similarity advantage. That is, there does not appear to be any redintegration effect with order accuracy measures a finding that is again consistent with most current models of memory where it is assumed that redintegration occurs only after order memory has been accessed.

The results of the current experiment conform to expectations in most respects. Age effects emerge and strong redintegration effects are present when item scoring is used, consistent with the notion that redintegration is occurring at the level of producing a candidate for recall. Likewise, redintegration effects are not apparent with order accuracy scoring. Knowing an item is a member of a particular category is unlikely to help a person identify at which position in the list the item occurred.

## Experiment 2

The explanation for performance in Experiment 1 is primarily a cueing argument which makes little reference to the underlying dimensions of the cue. That is, similarity is acting as a cue in order to facilitate the elicitation of potential candidates for recall. It is not a crucial

assumption that the items come from semantic categories. To test this assumption, in the following experiment we replicate the procedures of Experiment 1 but manipulate similarity by having participants study lists from rhyme categories or from non-rhyming sources. That is, phonological similarity is manipulated rather than semantic similarity.

In contrast to semantic similarity, one of the benchmark findings in immediate recall is the phonologically similar items are harder to recall, not easier, than phonologically dissimilar items. However, there are a growing number of instances where phonologically similar items are better recalled than dissimilar items (Fallon et al., 1999; Nairne & Kelley, 1999; Nimmo & Roodenrys, 2004). If our ideas are correct and we replicate the findings of Experiment 1, a linear relationship between task difficulty and similarity advantage should be apparent, particularly for item scoring. Given the robust findings that phonological similarity hurts order accuracy, the expectation would be that there would be no relationship between the size of the similarity effect and task difficulty in order accuracy but there should be an overall deficit in order accuracy for the similar lists given that the phonological representations of two rhyming items are similar to each other.

## Method

### *Participants*

Forty volunteers, either from the community or psychology undergraduates from the University of Southern Queensland, participated for tickets in the departmental prize draw or course credit. The sample of convenience consisted of 20 younger adults with an age range of 18 to 40 years ( $M = 27.05$ ,  $SD = 8.46$ ) and twenty older adults with an age range of 60 to 85 years ( $M = 69.15$ ,  $SD = 7.43$ ). All participants lived independently in the community, were native Australian English speakers, and had normal or correct-to-normal vision and hearing. Both participant groups reported themselves to be in good health and to have no difficulty reading the words as they would be presented on the computer screen. The mean number of

years of education did not differ between younger adults ( $M = 12.65$ ,  $SD = 2.11$ ) and older adults ( $M = 13.00$ ,  $SD = 3.58$ ),  $F(1,38) = 0.14$ . All participants were administered the Wechsler Test of Adult Reading (WTAR; Wechsler, 2001). The estimated IQ scores, based upon Wechsler Test of Adult Reading (WTAR; Wechsler, 2001) scores, were 108.75 and 114.60 for younger and older adults respectively. This difference was statistically significant,  $F(1,38) = 4.97$ .

### *Materials*

The method of list construction was identical to that used in Experiment 1. The word pool was generated by selecting six instances from each of 30 different rhyme categories from the South Florida Rhyme Category Norms (Walling, McEvoy, Oth, & Nelson, 1984). Stimuli were selected from medium sized rhyme categories (Mean number of items in the category was 21). All of the stimuli were one syllable words that shared the same rime but differed in their onsets (e.g. *lace, face, mace, race, chase, base*).

### *Procedure*

The procedure was identical to that used in Experiment 1.

## Results

### *Age-related effects in mean levels of performance*

The experimental design was a 2 (age) x 2 (similarity) x 2 (list length) x 2 (modality) x 3 (retention interval) mixed design, with age being the sole between-subjects variable. The table of means are presented in Appendix A. Robust benchmark effects of list length, modality, and retention interval effects were again readily apparent. However, in contrast to Experiment 1, there were no age effects in any of the three measures used.

### *Task Difficulty*

In the bottom right hand corner of Table 1 we present the rank order correlations (in italics) among the 12 estimates of task difficulty. As was the case in Experiment 1, the

ranking of our 12 estimates of task difficulty are very similar for similar and dissimilar lists, for younger and older participants for both correct-in-position and item scoring. Again the measures are not as good for order scoring.

### *Redintegration Effects*

#### *Correct-in-position*

As can be seen in the upper panel of Figure 3, with the easier conditions, there is a similarity decrement in that there is a negative similarity advantage. However, as task difficulty increases, a null similarity effect transitions to a positive similarity advantage. As was the case with Experiment 1, the strength of the relationship between task difficulty and the size of the similarity effect is quite low ( $r^2 = .38$  and  $.50$  for younger and older groups, respectively). There was no difference in the slopes for younger ( $b = .16$ ) and older participants ( $b = .12$ ),  $t(20) = .61$ ,  $p. > .05$ , nor for the intercepts ( $c = -.07$  and  $-.05$  for younger and older participants respectively),  $t(20) = .37$ ,  $p. > .05$ . The similarity of the slopes indicates that the underlying processes are much the same for younger and older adults.

-----  
 Insert Figure 3 about here  
 -----

#### *Item scoring*

The middle panel of Figure 3 depicts the redintegration effects for item scoring where the errors that determine task difficulty are the sum of omissions and any form of intrusion error in the dissimilar lists. For both younger and older participants there is a very strong relationship between task difficulty and the size of the similarity effect ( $r^2 = .91$  for younger adults and  $r^2 = .90$  for older adults). As task difficulty increases, from a slightly positive baseline, the size of the similarity advantage increases in a highly predictable way. The slopes for older ( $b = .39$ ) and younger participants ( $b=.32$ ) were not significantly different from each

other,  $t(20) = 1.38, p. > .05$ ; neither did the intercepts ( $c = .03$  and  $.06$ ),  $t(20) = 1.06, p. > .05$ .

### *Order accuracy*

As can be seen in the lower panel of Figure 3, the relationship between task difficulty and the similarity effect for order information is extremely weak ( $r^2 = .15$  and  $.01$  for young and old respectively). The regression line appears to be generally flat across all levels of task difficulty but centred at a 15% disadvantage for the similar items. The slope and intercepts are equivalent for young ( $b = .10, c = -.15$ ) and older ( $b = .03, c = -.15$ ) participants,  $t(20) = .59, p. > .05$  and  $t(20) = .06, p. > .05$ .

## Discussion

The results of the current experiment replicate those of Experiment 1 in all respects save two. There are no age differences in any of the measures used and for the order accuracy measure there is a similarity disadvantage at all levels of task difficulty.

The similarity results replicate previous findings in all respects. For the correct in position measure, there is a similarity disadvantage at low levels of task difficulty which reverse to a similarity advantage at high levels of task difficulty (Nairne & Kelley, 1999). When correct-in-position scoring is decomposed into its components, there is a consistent similarity advantage for item scoring and a consistent similarity decrement for order accuracy (Fallon et al., 1999).

With respect to the redintegration issue, again the data suggest that as task difficulty increases redintegration effects become more apparent. The results also suggest that the redintegration function is equivalent for younger and older participants.

One interesting aspect of the current results is that for item scoring, the regression equations for semantically similarity in Experiment 1 and for phonological similarity in Experiment 2 appear to be quite similar. As a follow up analysis, we directly compared the

similarity advantage across levels of task difficulty for semantic and phonological similarity. The data are depicted in Figure 4. The similarity advantage is equivalent for both types of similarity. There was no difference between the slopes ( $b = .36$  for phonemic and  $b = .38$  for semantic),  $t(44) = .41, p. > .05$ , nor between intercepts ( $c = .04$  for phonemic and  $c = .02$  for semantic),  $t(44) = 1.25, p. > .05$ .

-----  
Insert Figure 4 about here  
-----

### General Discussion

The current research explored memory performance of younger and older participants under task conditions where processing requirements are thought to be comparatively low and where, historically, age differences are harder to detect. Performance was evaluated within a redintegration framework (Schweickert, 1993) in which it is assumed that long-term lexical/semantic knowledge can be used to reconstruct a degraded phonological memory trace.

Short-term memory benchmark effects were readily apparent in the data, in that modality, list length, and retention interval effects were present for younger and older adults alike. Similarity effects also conform to prior findings. Thus, with semantic similarity there is a similarity advantage when item scoring is used, and no effect when order accuracy is measured (Saint-Aubin & Poirier, 1999b). Likewise, with phonological similarity, with correct-in-position scoring there is a similarity decrement under easy levels of task difficulty, which reverses to a similarity advantage when the task becomes more difficult (Fallon et al., 1999).

#### *Task Difficulty*

The redintegration account assumes that as task difficulty increases, the chances of direct retrieval become increasingly smaller and that back-up processes must be called into

play. In order to test this notion we manipulated modality of presentation, list length, and retention interval. Table 1 indicates that these effects combine in a way that is remarkably consistent across levels of similarity, across age groups, and even across experiments. In short, we have demonstrated a very reliable way of operationalising task difficulty.

### *Redintegration*

The results add to the literature in a number of ways. At the empirical level for both phonological and semantic similarity the size of the similarity advantage was highly predictable. That is, given knowledge about average performance on the dissimilar lists, it is possible to predict performance on the similar lists with some precision.

Finding a strong relationship between task difficulty and the size of the similarity advantage with semantic similarity, and then with phonological similarity, shows that this phenomenon is highly replicable. Moreover, the fact that the regression equations are equivalent for semantic and phonological similarity indicates that the similarity effects that we are observing are independent of the codes being employed. As such the data give compelling support for the redintegration perspective. Firstly, as task difficulty increases the memory trace, presumably phonological in nature, loses its fidelity. Then, as Saint-Aubin and Poirier (1999a, 1999b) suggest, similarity functions as a cue which points to a specific portion of LTM thereby enhancing the accessibility of potential candidates for recovery of the memory trace. In their terms, similarity narrows the search set in LTM. Note that the cuing function is again code independent as is implied in many accounts of short-term similarity effects (Fallon et al., 1999; Nairne & Kelley, 1999; Nimmo & Roodenrys, 2004; Poirier & Saint-Aubin, 1995, Saint-Aubin & Poirier, 1999b).

The role of similarity in redintegration clearly involves recovery of an item and says little, if anything, about where that item may have appeared in the list. The absence of any

systematic relationship between task difficulty and order accuracy again fits nicely with the redintegration framework.

Are the item effects that we have observed likely to generalise to other short-term phenomena? We think that this would only be expected in situations where it is plausible that a cue of some form is being used. Word frequency (Hulme et al., 1997) and concreteness/imagability (Walker & Hulme, 1999) effects have also been explained in terms of redintegration effects, but in these instances, the underlying mechanism is assumed to be differences in associative strength between phonological representations and their counterparts in lexical memory. We have no strong expectation that this form of redintegration would produce the same robust linear relationship with task difficulty, nor would we expect that regression equations would be equivalent.

### *Aging*

Age differences were present in Experiment 1, but not in Experiment 2. We have no firm explanation for this result and while some speculation is offered below, it the case that in short-term memory studies age differences are sometimes found and sometimes not. It is only meta-analytic studies (Bopp & Verhaeghen, 2005) that provide overall confirmation that there are age differences in absolute levels of recall in short-term memory tasks.

Modality, list length, and retention interval were manipulated to influence task difficulty at the general level, but were selected with the notion of testing various assumptions concerning short-term recall. Thus, auditory modality in most accounts is assumed to result in stronger or more discriminative registration in short-term memory. List length is aimed at taxing the capacity of any short-term store, and using a retention interval we hoped to control for rehearsal. While our analyses have not focused on these issues standard analysis of variance techniques performed on the means presented in Tables A1 and A2 indicate that no interactions involved age with any of the above variables. Consequently,



we would argue that any age differences in short-term recall are unlikely to be due to registration, storage capacity, or differential rates of forgetting. In this respect our results are consistent with much of the other aging short-term memory research (Kausler, 1994).

Our research addressed cognitive aging from the perspective of the Schweickert's (1993) notion of redintegration, a process that, in one form or another, is common to most current models of immediate memory. Our results suggest that the cueing or reduced search set version of redintegration is similar for both younger and older people. That is, both take advantage of the categorical nature of rhyming items or items from a taxonomic category to facilitate the reconstruction of a degraded candidate for recall.

To the best of our knowledge this is the first research that directly addresses aging effects in redintegration, but there is related work in the area. Oberauer (2001, 2005a, 2005b) has been exploring storage effects underpinning short-term memory/working memory within Cowan's (1995) "focus of attention" framework where a small number of items (and their episodic associations) can be maintained in a direct access region via the operation of the focus of attention. In addition, items (and their episodic associations) that have recently been studied are in the activated region in LTM in the sense that these items are at above baseline levels of long-term activation. Interestingly, his work has indicated that there are no age differences in the direct access component, but that age effects may reflect aspects of residual activation in the LTM component of the model. The current research compliments Oberauer's findings in that the direct access region provides a supporting mechanism for ensuring an undegraded trace that leads to direct and successful recall. Moreover, Oberauer (2005) argues that with the items in activated LTM, "if the activation of content representations one wishes to maintain is sufficiently distinct, it can serve to recover the identity of these contents."(p.727). Presumably, he envisages the representations of the

activated item in LTM items to be somewhat degraded but can be recovered via a redintegration like process.

An alternative general theory that has been postulated for aging effects is that as one grows older cognitive resources become depleted. Alternatively, it is possible that neural degeneration with age may produce memory traces that are more “noisy”. Such a perspective fits well with the redintegration approach in that increased levels of noise in a memory trace is equivalent to saying that the memory trace has less fidelity and as such redintegration is going to be required to augment recall. The noise account of aging thus readily accounts for the current data in that in all respects the recall process for younger and older participants is equivalent.

Finally, it has been recently suggested that age deficits in short-term memory tasks may not be memory deficits at all, but instead reflect perceptual processing deficits. Surprenant, Neath, and Brown (2006) examined the relationship between hearing ability and memory for auditorially presented lists of phonologically similar and dissimilar consonants. Using multiple dimensional scaling techniques they established that the similarity functions for older adults were more compressed than that of younger participants. That is, the representations of both phonologically similar and dissimilar items are more similar for older people than they are for young people. Surprenant et al. used these characteristics when modelling the age-related decrement in recall of phonologically similar and dissimilar lists with the SIMPLE (Brown, Neath, & Chater, in press; Neath & Brown, 2006) framework. The data and the model were able to account simultaneously for age, phonological similarity, serial position, and error effects. They argued that the age-related decrement in memory could be attributed in part to the fact that even slight problems in auditory acuity may produce memory traces that are less distinctive.

The research has demonstrated that when serial recall is decomposed into its component parts, similarity effects can be accurately predicted from some knowledge of task difficulty with item scoring but not for order scoring. The regression equation describing the relationship between similarity and task difficulty for item scoring is equivalent for older and younger participants and for phonological and semantic similarity. These findings provide compelling evidence for a cuing function within a redintegration framework, and suggest that the emergence of age differences in short-term tasks may well be due to decreased fidelity of the short-term memory trace that increases the necessity for redintegration.

## References

- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press
- Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *Journals of Gerontology Series B*, *60*, 223-233.
- Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory & Language*, *34*, 594-621.
- Brown, G. D. A., Neath, I., & Chater, N. (In press). A temporal ration model of memory. *Psychological Review*.
- Burgess, N., & Hitch, G. J. (1996). A connectionist model of STM for serial order. In S. E. Gathercole (Ed.), *Models of short-term memory*, (pp. 51-71). Hove, UK: Psychology Press.
- Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. A. (2006). The real relationship between short-term memory and working memory. *Memory*, *14*, 804-813.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake and P. Shah (Eds.) *Models of Working Memory: Mechanisms of active maintenance and executive control*. (pp. 102-134). New York: Cambridge University Press.
- Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp.191-211). New York: Plenum.
- Engle, R. W., Kane, M. J. & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake and P. Shah (Eds.)

- Models of Working Memory: Mechanisms of active maintenance and executive control.* (pp. 102-134). New York: Cambridge University Press.
- Fallon, A. B., Groves, K., & Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT, but not MAN. *International Journal of Psychology, 34*, 301-307.
- Hasher, L., & Zacks, H. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193-225). San Diego, CA: Academic Press.
- Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology, 36*, 73-137.
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, M., & Stuart, G. (1997). Word frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1217-1232.
- Kausler, D. H. (1994). *Learning and memory in normal aging*. San Diego: Academic Press.
- Kramer, A. F., Hahn, S., & Gopher, D. (1999). Task coordination and aging: Explorations of executive control processes in the task switching paradigm. *Acta Psychologica, 101*, 339-378.
- Mayr, U., Spieler, D. H., & Kliegel, R. (2001). *Aging and executive control*. Hove, UK: psychology Press.
- McEvoy, C. L., & Nelson, D. L. (1982). Category name and instance norms for 106 categories of various sizes. *American Journal of Psychology, 95*, 581-634.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251-269.

- Nairne, J., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & cognition, 27*, 45-53.
- Neath, I., & Brown, G. D. A. (2006). SIMPLE: Further applications of a local distinctiveness model of memory. In B. H. Ross (Ed.) *The psychology of learning and motivation*. (pp. 201-243). San Diego, CA: Academic Press.
- Nimmo, L. M., & Roodenrys, S. (2004). Investigating the phonological similarity effect: Syllable structure and the position of common phonemes. *Journal of Memory and Language, 50*, 245-258.
- Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 948-957.
- Oberauer, K. (2005a). Control of the contents of working memory – a comparison of two paradigms and two age groups. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 714-728.
- Oberauer, K. (2005b). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General, 134*, 368-387.
- Oberauer, K., & Kliegl, R. (2001). Beyond resources: Formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology, 13*, 187-215.
- Page, M.P.A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761-781.
- Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence of the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology, 48A*, 384-404.

- Rouleau, N., & Belleville, S. (1996). Irrelevant speech effect in aging: An assessment of inhibitory processes in working memory. *Journal of Gerontology: Psychological Sciences, 51B*, 356-363.
- Saint-Aubin, J., & Poirier, M. (1999a). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology, 34*(5/6), 347-352.
- Saint-Aubin, J., & Poirier, M. (1999b). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *The Quarterly Journal of Experimental Psychology, 52A*, 367-394.
- Salthouse, T. A. (1996). Processing-speed theory of adult age differences in cognition. *Psychological review, 103*, 403-428.
- Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory and Cognition, 21*, 168-175.
- Schweickert, R., Chen, S., & Poirier, M. (1999). Redintegration and the useful lifetime of the verbal memory representation. *International Journal of Psychology, 34*, 447-453.
- Surprenant, A. M., Neath, I., & Brown, G. D. A. (in press). Modeling age-related differences in immediate memory using SIMPLE. *Journal of Memory and Language, 55*, 572-586.
- Tehan, G., & Humphreys, M. S. (1995). Transient phonemic codes and immunity to proactive interference. *Memory and Cognition, 23*, 181-191.
- Tolan, G. A. & Tehan, G. (1999). Determinants of Short-term Forgetting: Decay, Retroactive Interference or Proactive Interference? *International Journal of Psychology, 34*, 285-292.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words:

Evidence for a semantic contribution to short-term serial recall. *Journal of*

*Experimental Psychology: Learning, Memory and Cognition*, 25, 1256-1271.

Walling, J. R., McEvoy, C. L., Oth, J. E., & Nelson, D. L. (1984). Rhyme category norms.

Unpublished manuscript. The University of South Florida.

Wechsler, D. (2001). *WTAR technical manual*. San Antonio, TX: The Psychological

Corporation.



Table 1

Rank order correlations

	Semantic Young	Semantic Young	Semantic Older	Semantic Older	Phonemic Young	Phonemic Young	Phonemic Older	Phonemic Older
<b>Correct In Position</b>	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar	Similar	Dissimilar
Semantic Young								
Similar	1							
Semantic Young								
Dissimilar	<b>0.99</b>	1						
Semantic Older								
Similar	<b>0.99</b>	<b>0.97</b>	1					
Semantic Older								
Dissimilar	<b>0.99</b>	<b>0.99</b>	<b>0.98</b>	1				
Phonemic Young								
Similar	0.93	0.91	0.97	0.91	1			
Phonemic Young								
Dissimilar	0.99	0.98	0.98	0.97	0.94	1		

Phonemic Older

Similar 0.99 0.97 1 0.98 0.97 0.98 1

Phonemic Older

Dissimilar 1 0.99 0.99 0.99 0.93 0.99 0.99 1

**Item Scoring**

Semantic Young

Similar 1

Semantic Young

Dissimilar **0.97** 1

Semantic Older

Similar **0.99** **0.98** 1

Semantic Older

Dissimilar **0.97** **0.99** **0.97** 1

Phonemic Young

Similar 0.95 0.99 0.95 0.99 1

Phonemic Young

Dissimilar 0.94 0.99 0.96 0.99 0.99 1

Phonemic Older

Similar 0.97 0.99 0.97 1 0.99 0.99 1

Phonemic Older

Dissimilar 0.93 0.98 0.95 0.97 0.98 0.99 0.97 1

**Order Scoring**

Semantic Young

Similar 1

Semantic Young

Dissimilar **0.97** 1

Semantic Older

Similar **0.99** **0.98** 1

Semantic Older

Dissimilar **0.97** **0.99** **0.97** 1

Phonemic Young

Similar 0.63 0.69 0.68 0.65 1

Phonemic Young

Dissimilar 0.78 0.85 0.83 0.8 0.94 1

Phonemic Older

Similar	0.79	0.85	0.84	0.81	<i>0.94</i>	<i>0.99</i>	1
---------	------	------	------	------	-------------	-------------	---

Phonemic Older

Dissimilar	0.92	0.9	0.94	0.88	<i>0.78</i>	<i>0.87</i>	<i>0.88</i>	1
------------	------	-----	------	------	-------------	-------------	-------------	---

### Acknowledgements

The authors would like to thank the older participants from the Toowoomba District who volunteered to participate in the research. Our thanks also go to Drs Ian Neath, Jean Saint-Aubin and two anonymous reviewers for constructive comments on an earlier draft of the manuscript. This research was supported by a USQ post-graduate scholarship to the first author.

## Appendix A

Mean levels (and standard error of the mean) of performance for Experiment 1 are presented in Table A1 and for Experiment 2 are presented in Table A2. (Note outcomes of 2x2x2x2x3 ANOVA can be retrieved from <http://www.usq.edu.au/users/tehan/ageanova.doc>.)

Table A1. Mean recall (standard error of mean) for semantically similar and dissimilar lists as a function of age, modality, list length, and retention interval.

			Correct in Position								
			Younger				Older				
			Similar		Dissimilar		Similar		Dissimilar		
Read Silent	4-Item	Immediate	0.89	0.03	0.81	0.04	0.83	0.04	0.74	0.04	
		2-Sec	0.64	0.03	0.49	0.04	0.52	0.04	0.36	0.04	
		4-Sec	0.58	0.04	0.39	0.03	0.46	0.04	0.28	0.04	
	6-Item	Immediate	0.49	0.03	0.39	0.03	0.40	0.03	0.26	0.02	
		2-Sec	0.38	0.02	0.26	0.02	0.30	0.03	0.20	0.02	
		4-Sec	0.36	0.02	0.20	0.02	0.25	0.03	0.15	0.02	
	Read Aloud	4-Item	Immediate	0.97	0.02	0.95	0.02	0.92	0.02	0.91	0.03
			2-Sec	0.82	0.02	0.75	0.04	0.72	0.04	0.62	0.04
			4-Sec	0.71	0.03	0.64	0.05	0.63	0.04	0.51	0.04
6-Item		Immediate	0.49	0.03	0.50	0.04	0.43	0.03	0.36	0.04	
		2-Sec	0.38	0.02	0.38	0.04	0.37	0.03	0.27	0.02	
		4-Sec	0.36	0.02	0.29	0.02	0.26	0.02	0.21	0.03	
			Item Scoring								
Read Silent	4-Item	Immediate	0.93	0.02	0.87	0.03	0.90	0.02	0.80	0.04	
		2-Sec	0.79	0.02	0.64	0.04	0.74	0.03	0.52	0.03	
		4-Sec	0.74	0.03	0.52	0.03	0.70	0.03	0.41	0.03	
	6-Item	Immediate	0.72	0.03	0.57	0.03	0.66	0.03	0.47	0.03	
		2-Sec	0.66	0.03	0.42	0.03	0.56	0.02	0.35	0.02	
		4-Sec	0.61	0.02	0.37	0.03	0.56	0.02	0.31	0.01	
Read Aloud	4-Item	Immediate	0.98	0.01	0.98	0.01	0.97	0.01	0.95	0.01	
		2-Sec	0.93	0.02	0.86	0.02	0.88	0.02	0.78	0.03	
		4-Sec	0.90	0.02	0.76	0.03	0.84	0.03	0.68	0.03	
	6-Item	Immediate	0.86	0.02	0.68	0.03	0.78	0.02	0.62	0.03	

2-Sec	0.77	0.02	0.59	0.03	0.70	0.02	0.49	0.02
4-Sec	0.76	0.02	0.47	0.03	0.67	0.02	0.41	0.02

Order Accuracy

Read Silent	4-Item	Immediate	0.95	0.02	0.92	0.02	0.92	0.02	0.92	0.02
		2-Sec	0.82	0.04	0.77	0.04	0.70	0.04	0.68	0.04
		4-Sec	0.78	0.03	0.74	0.04	0.65	0.04	0.67	0.06
	6-Item	Immediate	0.69	0.03	0.67	0.04	0.60	0.04	0.55	0.04
		2-Sec	0.58	0.03	0.64	0.04	0.54	0.04	0.57	0.04
		4-Sec	0.59	0.02	0.55	0.04	0.45	0.05	0.47	0.05
Read Aloud	4-Item	Immediate	0.98	0.01	0.97	0.01	0.95	0.02	0.96	0.02
		2-Sec	0.88	0.02	0.86	0.03	0.81	0.04	0.79	0.03
		4-Sec	0.79	0.03	0.82	0.04	0.75	0.04	0.73	0.04
	6-Item	Immediate	0.72	0.04	0.73	0.05	0.55	0.04	0.57	0.05
		2-Sec	0.58	0.04	0.64	0.04	0.52	0.04	0.57	0.04
		4-Sec	0.56	0.03	0.60	0.03	0.39	0.03	0.50	0.05



Table A2. Mean recall (standard error of mean) for phonologically similar and dissimilar lists as a function of age, modality, list length, and retention interval.

			Correct in Position							
			Younger				Older			
			Rhyming		Dissimilar		Rhyming		Dissimilar	
Read Silent	4-Item	Immediate	0.70	0.04	0.73	0.03	0.71	0.04	0.71	0.06
		2-Sec	0.41	0.04	0.39	0.04	0.41	0.05	0.40	0.05
		4-Sec	0.41	0.04	0.30	0.03	0.35	0.04	0.34	0.04
	6-Item	Immediate	0.34	0.04	0.31	0.04	0.27	0.02	0.25	0.04
		2-Sec	0.26	0.03	0.21	0.03	0.22	0.03	0.16	0.03
		4-Sec	0.25	0.03	0.17	0.03	0.21	0.03	0.12	0.02
Read Aloud	4-Item	Immediate	0.81	0.04	0.87	0.03	0.82	0.03	0.84	0.03
		2-Sec	0.57	0.04	0.53	0.05	0.54	0.05	0.60	0.06
		4-Sec	0.52	0.04	0.46	0.04	0.47	0.04	0.45	0.04
	6-Item	Immediate	0.30	0.03	0.39	0.05	0.32	0.03	0.34	0.05
		2-Sec	0.25	0.03	0.22	0.04	0.25	0.03	0.22	0.03
		4-Sec	0.24	0.03	0.18	0.03	0.22	0.02	0.18	0.03
			Item Scoring							
Read Silent	4-Item	Immediate	0.94	0.01	0.82	0.03	0.90	0.02	0.80	0.04
		2-Sec	0.73	0.02	0.54	0.03	0.74	0.03	0.53	0.04
		4-Sec	0.66	0.04	0.42	0.02	0.67	0.03	0.48	0.04
	6-Item	Immediate	0.73	0.02	0.52	0.03	0.71	0.02	0.50	0.03
		2-Sec	0.61	0.02	0.37	0.03	0.59	0.03	0.31	0.03
		4-Sec	0.59	0.03	0.30	0.02	0.58	0.03	0.26	0.02
Read Aloud	4-Item	Immediate	0.98	0.01	0.90	0.02	0.96	0.01	0.88	0.02
		2-Sec	0.88	0.01	0.68	0.04	0.85	0.03	0.69	0.05
		4-Sec	0.81	0.02	0.63	0.03	0.82	0.02	0.63	0.03
	6-Item	Immediate	0.76	0.02	0.60	0.03	0.78	0.02	0.55	0.04

2-Sec	0.71	0.02	0.46	0.03	0.72	0.03	0.43	0.03
4-Sec	0.67	0.02	0.39	0.02	0.68	0.02	0.35	0.04

Order Accuracy

Read Silent	4-Item	Immediate	0.74	0.04	0.89	0.02	0.79	0.03	0.86	0.05
		2-Sec	0.55	0.05	0.70	0.05	0.54	0.05	0.72	0.05
		4-Sec	0.60	0.04	0.68	0.04	0.50	0.05	0.68	0.05
	6-Item	Immediate	0.48	0.05	0.57	0.06	0.38	0.03	0.47	0.06
		2-Sec	0.41	0.04	0.50	0.06	0.37	0.05	0.46	0.08
		4-Sec	0.41	0.04	0.54	0.06	0.35	0.04	0.44	0.06
Read Aloud	4-Item	Immediate	0.82	0.03	0.95	0.01	0.86	0.02	0.95	0.02
		2-Sec	0.64	0.04	0.75	0.05	0.61	0.05	0.84	0.03
		4-Sec	0.63	0.04	0.72	0.05	0.57	0.04	0.70	0.05
	6-Item	Immediate	0.40	0.04	0.62	0.05	0.41	0.03	0.59	0.04
		2-Sec	0.34	0.04	0.44	0.05	0.34	0.03	0.49	0.05
		4-Sec	0.35	0.05	0.41	0.05	0.32	0.03	0.52	0.05

Figure Captions:

Figure 1. Results of a hypothetical study presented in standard format (upper panel) or in ranked order (lower panel).

Figure 2. Semantic similarity advantage as a function of task difficulty for correct-in-position (upper panel), item scoring (middle panel) and order accuracy scoring (lower panel) for younger (diamonds) and older (squares) participants.

Figure 3. Semantic similarity advantage as a function of task difficulty for correct-in-position (upper panel), item scoring (middle panel) and order accuracy scoring (lower panel) for younger (diamonds) and older (squares) participants.

Figure 4. Combined similarity advantage for item scoring as a function of task difficulty.

Figure 1

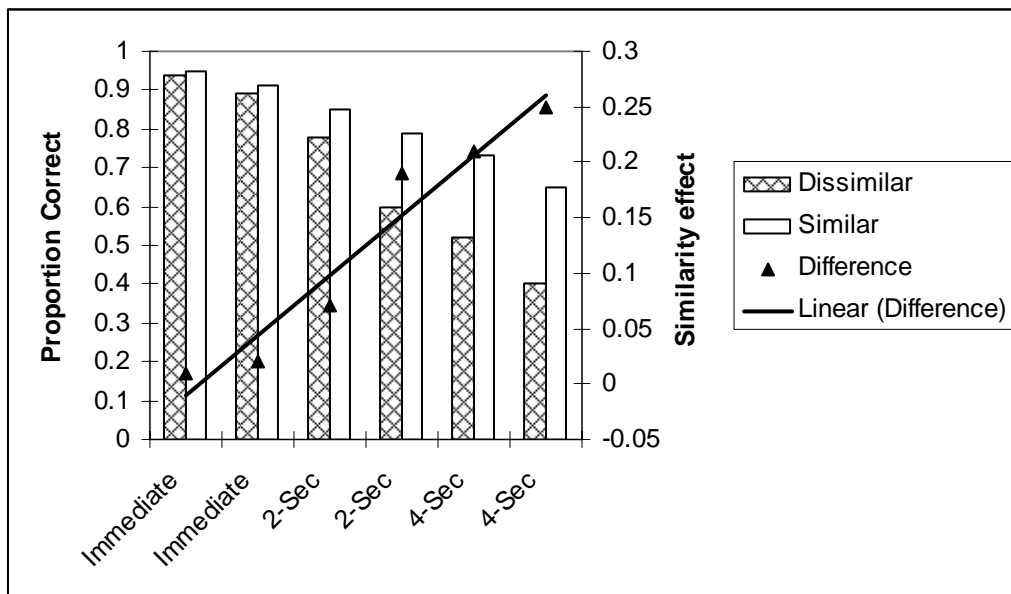
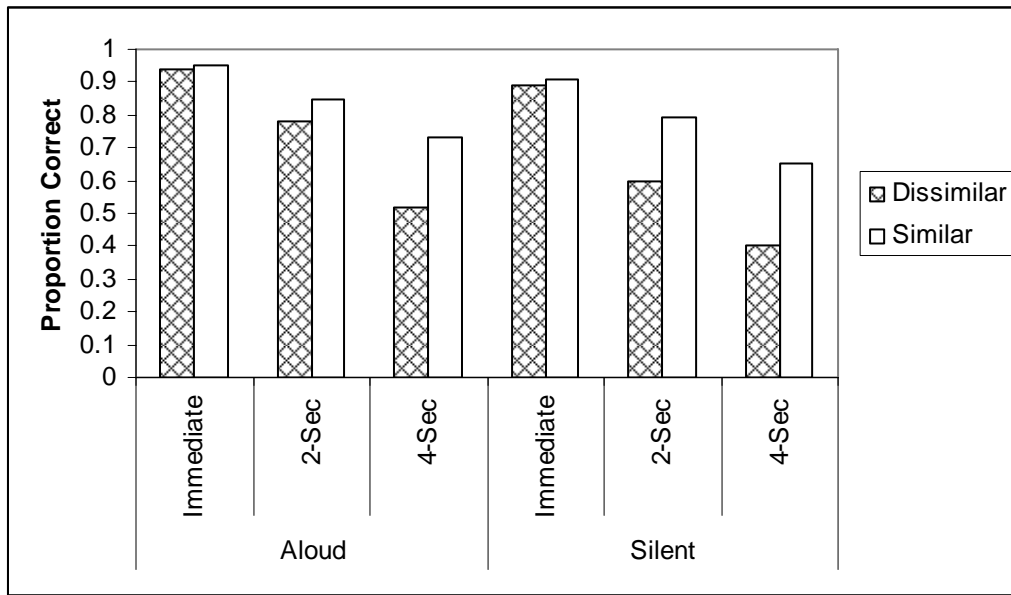


Figure 2.

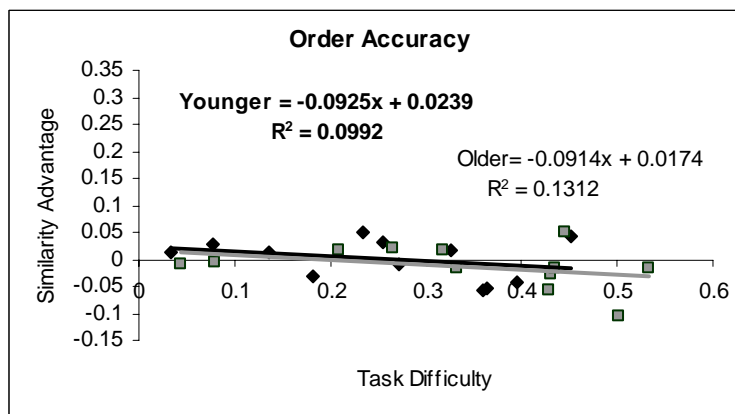
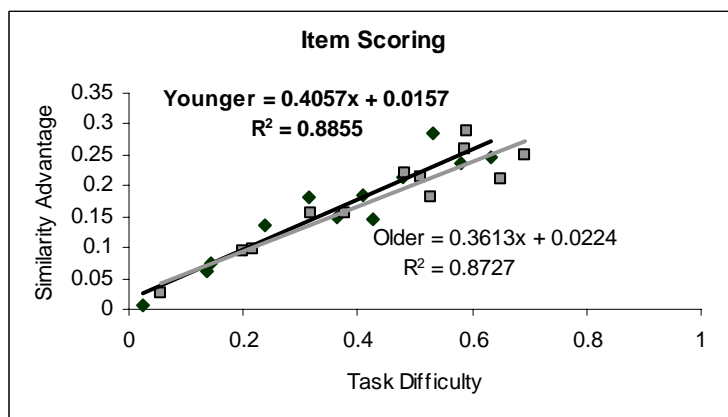
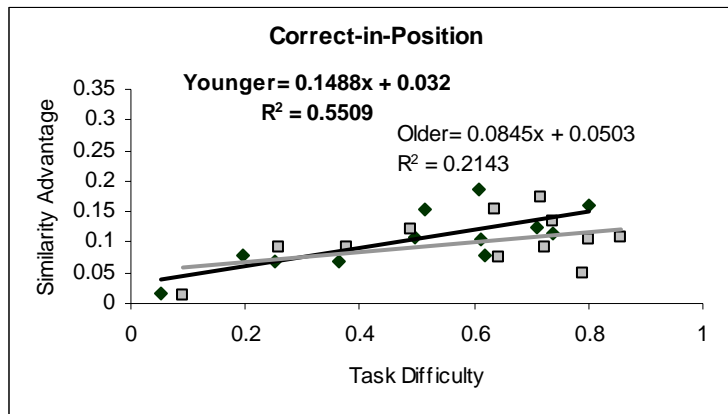


Figure 3.

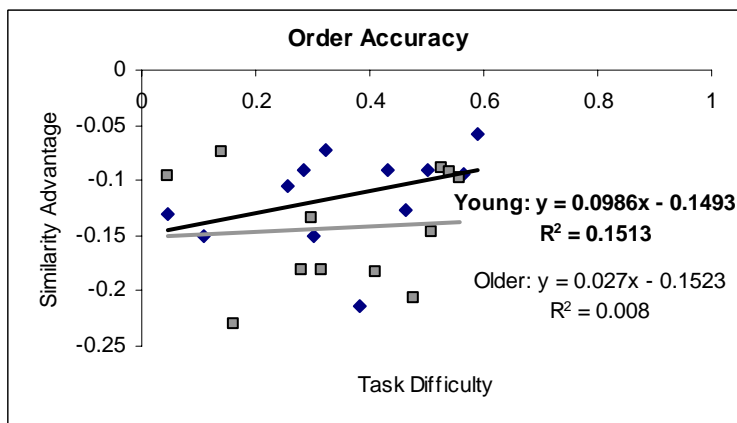
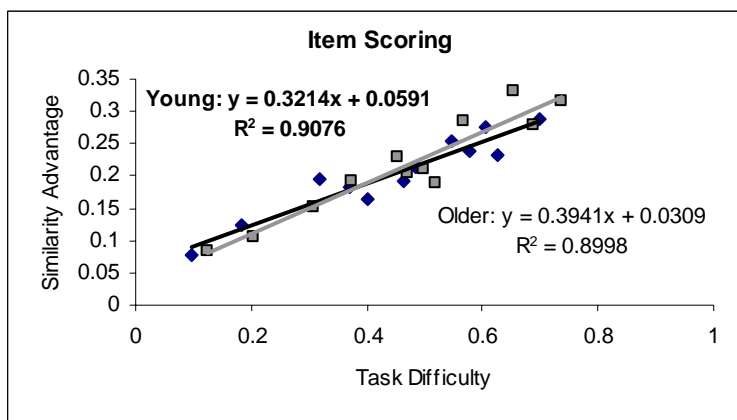
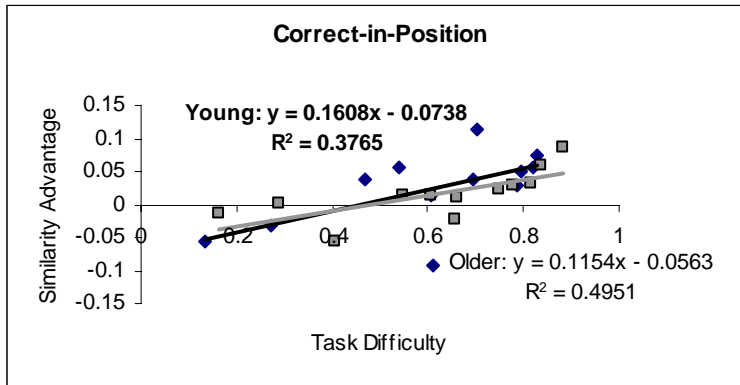


Figure 4.

