Emery, L., Hale, S., & Myerson, J. (2008). Age differences in proactive interference, working memory, and reasoning. **Psychology and Aging**, 23(3), 634-45. (Sep 2008) Published by the American Psychological Association (ISSN: 1939-1498). DOI: 10.1037/

Age Differences in Proactive Interference, Working Memory, and Abstract Reasoning

Lisa Emery, Sandra Hale, and Joel Myerson

ABSTRACT

It has been hypothesized that older adults are especially susceptible to proactive interference (PI) and that this may contribute to age differences in working memory performance. In young adults, individual differences in PI affect both working memory and reasoning ability, but the relations between PI, working memory, and reasoning in older adults have not been examined. In the current study, young, old, and very old adults performed a modified operation span task that induced several cycles of PI buildup and release as well as two tests of abstract reasoning ability. Age differences in working memory scores increased as PI built up, consistent with the hypothesis that older adults are more susceptible to PI, but both young and older adults showed complete release from PI. Young adults' reasoning ability was best predicted by working memory performance under high PI conditions, replicating M. Bunting (2006). In contrast, older adults' reasoning ability was best predicted by their working memory performance under low PI conditions, thereby raising questions regarding the general role of susceptibility to PI in differences in higher cognitive function among older adults.

The working memory system is often described as a mental workspace used for the temporary maintenance and manipulation of information in the service of thought (Baddeley, 1986). The functioning of the working memory system is typically measured using working memory span tests, in which a participant must remember a short list of items while performing mental manipulation, transformation, or calculation. Memory span tasks such as reading span (Daneman & Carpenter, 1980) and operation span (Turner & Engle, 1989) are validated by their correlations with performance on higher order cognitive tasks that are supported by the working memory system, such as abstract reasoning (e.g., Engle, Tuholski, Laughlin, & Conway, 1999) and language comprehension (e.g., Daneman & Merikle, 1996). Consequently, anything that decreases the efficiency of the working memory system should be reflected in lower working memory span scores, and if the portion of the working memory system affected is important for thought, then this decrease in efficiency should also cause poorer performance on higher order cognitive tasks.

Age-related variation in the working memory system is reflected in both cross-sectional age differences (e.g., Hale, Myerson, Emery, Lawrence, & Dufault, 2007; Park, Lautenschlager, Hedden, Davidson, & Smith, 2002; Salthouse, 1993) and longitudinal age changes (e.g., Hertzog, Dixon, Hultsch, & MacDonald, 2003; Schaie, 2005) in both working memory spans and abstract reasoning tasks. Explanations of this age-related variation have focused on the role of processing speed (Salthouse, 1994) and executive function, broadly defined (Hedden & Yoon, 2006; Myerson, Emery, White, & Hale, 2003) as possible causes of age differences on working memory and higher order tasks. Within the realm of executive function, there has been increasing interest in the role that the ability to resist proactive interference (PI) may play in determining age differences in working memory spans (Bowles & Salthouse, 2003; May, Hasher, & Kane, 1999) and whether the ability to resist PI is related to individual differences in higher order cognition in younger (Bunting, 2006) and older adults (Lustig, Hasher, & May, 2001).

PI occurs when information that has previously been remembered interferes with memory for new information. May et al. (1999) pointed out that working memory span tasks present repeated trials of material to be remembered, recalled, and then forgotten and thus represent an ideal situation for the buildup of PI. May et al. further suggested that the traditional method of administering trials from the lowest number of words to the highest number of words may penalize individuals (i.e., older adults) who have a particular problem with PI, because the trials with the longest series are those in which PI may be the highest. In order to test this hypothesis, May et al. (1999) gave older and younger adults either a traditional ascending version of the reading span test or a descending version in which the more difficult trials (e.g., trials with more words to remember) were administered first. Typical age differences were found on the ascending version but not on the descending version, supporting the hypothesis of decreased ability to resist PI in older adults.

The working memory system is multifaceted, however, and even though a factor influences working memory span scores, it may not tap those aspects of the working memory system that are related to higher order cognition. To see if the ability to resist PI was related to higher order cognition, Lustig et al. (2001) replicated the May et al. (1999) finding and further found that in older adults only the ascending (high PI) version of the reading span task predicted individual differences in prose recall.

One limitation of the ascending versus descending approach is that PI is not directly measured in these tasks but is inferred from the results. In addition, there is the possibility that other task differences may contribute to the pattern of performance across age groups. For example, the ascending and descending procedures may differ on perceived difficulty. Because older adults may show more stress reactivity in cognitive testing situations than younger adults (Neupert, Soederberg, & Lachman, 2006), this could affect older adults' performance more than it affects the younger adults' performance. One way of solving both these problems is to use a working memory test in which PI is directly manipulated and measured across trials in order to see whether age differences in span scores vary as the amount of PI varies and whether the relationship between working memory span and higher cognitive ability varies as the amount of PI varies.

One such test has recently been used to investigate individual differences in PI resistance in younger adults (Bunting, 2006). In Bunting's (2006) study, young adult participants performed a version of the operation span working memory task in which a release-from-PI test (Wickens, 1970) was embedded. Participants performed 12 operation span trials, in which the memory material switched between words and numbers every 3 trials. The switch trials (low PI) served to release PI that was built up over the previous trials (moderate and high PI). Bunting (2006) found that the moderate and high PI trials were more predictive of performance on the Ravens Advanced Progressive Matrices (RAPM) than were the low PI trials, consistent with the hypothesis that the ability to resist PI is an important component of individual differences in working memory ability.

The current study had two goals. The first was to test the hypothesis that age differences in susceptibility to PI are at least partly responsible for age differences in working memory performance using a version of Bunting's (2006) release-from-PI operation span task. This design avoids the limitations of the ascending versus descending span comparison by directly manipulating and measuring the amount of PI involved in the working memory task. If older adults are particularly susceptible to PI and this contributes to their poor performance on working memory tasks, then age differences should be largest on those trials in which PI is greatest. The second goal of this study was to determine whether varying the amount of PI influences the correlation between working memory and reasoning ability in older adults in the same way that it does in young adults (Bunting, 2006). The fundamental question here is the extent to which older adults' increased susceptibility to PI affects their ability to reason.

METHOD

Participants

Participants were recruited in three age groups: young adults (ages 18–29 years, n = 71), older adults (ages 60–79 years, n = 71), and very old adults (ages 80–95 years, n = 20). Participants were primarily recruited through the Volunteers for Health pool maintained by the Washington University School of Medicine. Additional participants were recruited from the older adult and undergraduate participant pools maintained by the Washington University Department of Psychology. All participants (except for the Washington University undergraduates, who received course credit) were paid $10.00 per hour for participation.

Prior to participation, participants were screened for a history of neuropsychological (e.g., stroke, dementia) or serious psychiatric (e.g., schizophrenia, bipolar disorder) problems. During the testing session, participants completed another longer health screening questionnaire, the short form of the Geriatric Depression Scale (Sheikh & Yesavage, 1986) and the Short Blessed Test of Memory and Orientation (Katzman et al., 1983). Six participants (3 young adults, 2 older adults, and 1 very old adult) were excluded from analysis as a result of this screening. In addition, the data from 1 young adult who did not perform the release-from-PI operation span properly (less than 85% correct on the math portion) and 1 older adult who did not complete all of the tasks were excluded from the analyses as well as the data from 1 older adult who was an influential outlier in several cases (i.e., DFBetas greater than 2.0). This left totals of 67 young adults, 67 older adults, and 19 very old adults in the analyses reported below.

Release-From-PI Operation Span

The release-from-PI operation span procedure was modified from Bunting (2006) in two ways. First, rather than using digits and words to provide PI buildup and release, the current test used various taxonomic categories of words. This was done to eliminate potential difficulties in interpretation caused by known larger age differences for word than for digit memory (Verhaeghen, Marcoen, & Goossens, 1993). Second, rather than using series of all the same length (six items), participants performed the release-from-PI Operation Span at series lengths of two, three, and four words. Previous studies examining working memory and age differences in PI resistance used series lengths of two to four words (Lustig et al., 2001; May et al., 1999), facilitating comparison with the current study. In addition, previous researchers have hypothesized that age differences in memory span tasks, as typically administered, are due to age differences in the ability to resist PI. Because six items is well beyond the operation span of most young and (particularly) older adults, performance on trials of only this length would not provide direct evidence for this hypothesis. Finally, recent evidence has suggested that the predictive validity of complex span measures like operation span is roughly the same across items of different list lengths, including the range used here (Unsworth & Engle, 2006). Thus, sufficient correlations between reasoning ability and operation span performance should be detected even at these shorter list lengths.

Structure and Procedure

On each trial of the release-from-PI operation span task, participants saw a set of equation–word pairs presented one by one on the computer screen (e.g., “IS (2 + 4) – 1 = 5? COW”). As soon as each equation–word pair was presented, participants read the equation aloud and judged whether the equation was true. Participants then read the word aloud. To prevent extra rehearsal of the words, presentation rate was controlled by the experimenter, who advanced to the next equation as soon as the word was read aloud (Friedman & Miyake, 2004). After each set of sentences, the word RECALL appeared on the screen and participants were instructed to recall the memory words for the current set.

Participants were given 12 trials at each of three list lengths (two, three, or four words), for a total of 36 trials. Each set of 12 trials included four cycles of PI buildup—that is, four successive taxonomic categories of words—with 3 trials in each cycle. The list lengths were presented in ascending order—that is, participants were first given all 12 trials of two-word lists, then all 12 trials of three-word lists, and finally all 12 trials of the four-word lists. To eliminate the buildup of PI between series lengths, participants were given short, filled breaks in which they performed a visual processing speed task as a filler task. Each break lasted a minimum of 5 min. During the first break, participants completed the Weschler Adult Intelligence Scale–Third Edition (WAIS-III) Digit-Symbol Substitution test (Psychological Corporation, 1997); during the second break, participants completed the WAIS-III Symbol Search test.

Stimuli

The words used in the release-from-PI operation span task were chosen from the Battig and Montague (1969) norms. All words were one or two syllables in length and were selected to be as phonologically dissimilar as possible given the constraints of category membership. In addition, data from the English Lexicon Project (Balota et al., 2002) were used to ensure that the words chosen had similar naming times, lexical decision times, and word frequency across relevant variables. That is, these three variables were equivalent across list length (two vs. three vs. four words), buildup cycle (first vs. second vs. third vs. fourth), and PI level (low vs. moderate vs. high). In addition, an analysis of variance (ANOVA) conducted for each dependent variable (naming time, lexical decision time, and word frequency) indicated that there were no significant interaction effects of any of the independent variables (list length, buildup cycle, or PI level). [1]

The equations used in the release-from-PI operation span task involved only addition and subtraction, with each correct solution between 2 and 9. Incorrect solutions were generated by randomly adding or subtracting 2 from the correct answer. Half of the equations presented were correct, and half were incorrect. The stimuli, both words and equations, used in the release-from-PI operation span task are presented in the Appendix.

Scoring

Following Bunting (2006), participants were eliminated from the analysis if they scored lower than 85% correct on the math portion of the operation span task. Trials were then scored as correct (1 point) or incorrect (0 points) depending on whether the participant correctly recalled all of the words in the trial (and no additional words).

Other Measures

Traditional Operation Span

Participants also performed a traditional version of the operation span task, similar to that used in previous studies of individual differences (e.g., Engle et al., 1999). In this version, list length ranged from two to five words, with 4 trials at each list length, for a total of 16 trials. The trials were presented in four blocks. Each block contained 1 trial from each of the four list lengths, and within each block the list lengths were randomly presented. As in the release-from-PI operation span task, the words for the traditional operation span task were selected from the Battig and Montague (1969) norms but were not arranged into categories. The word used in the traditional operation span task did not differ in frequency, naming time, or lexical decision time from those in the release-from-PI operation span task. Scoring was completed as in the release-from-PI operation span task, for a possible score of 16.

Wechsler Abbreviated Scale of Intelligence (WASI)

The two-test version of the WASI (Psychological Corporation, 1999) was administered to ensure individuals of different age groups were relatively similar in (age-normed) IQ and to provide a measure of abstract reasoning ability. The two-test version of the WASI includes shorter versions of the Vocabulary and Matrix Reasoning tests from the full WAIS-III and can provide an estimate of full-scale IQ. Administration of the Vocabulary test was done as instructed in the WASI manual. The administration of the Matrix Reasoning test was modified slightly as follows.

The standardized administration of the Matrix Reasoning sets different starting points for each age group tested, and administration of the test stops either when the participant reaches the stopping point for their age group or a discontinuation criterion (four incorrect responses out of the last five attempted), whichever comes first. For the current study, the appropriate starting point (i.e., as described in the WASI manual) was used for each age group in order to get accurate IQ estimates. However, only the discontinuation criterion was used for a stopping point. That is, if an individual had reached the appropriate stopping point for their age group but had not yet missed four out of the last five items, they were allowed to continue. IQs were then calculated on the basis of the stopping point for the appropriate age, but their raw score (used for most analyses) was based on their score using the stopping criterion.

RAPM

Participants also performed Sets I and II of RAPM. This test is similar to Matrix Reasoning but is designed to differentiate among people of “superior intellectual ability” (Raven, Raven, & Court, 1993). This test was chosen as the main measure of abstract reasoning because it is considered a relatively knowledge-free measure of reasoning ability and has been used previously in studies of both individual differences (i.e., Bunting, 2006) and age differences (i.e., Salthouse, 1993) in working memory and reasoning.

Administration of the RAPM was modified from the standard instructions as follows. First, problems were presented one by one, as they were presented in the Matrix Reasoning test. Participants gave their answer aloud before proceeding to the next problem. Participants were first given the 12-problem Set I and had to correctly answer at least 4 of the 12 problems (25%) to continue to Set II. For the 36 problems of Set II, testing was discontinued when a participant incorrectly answered 5 problems in a row or answered 5 incorrectly out of the last 6. Participants were given as much time as they needed to answer each problem.

Procedure

Participants were tested individually in sessions lasting between 2 and 3 hr. Upon arrival, participants filled out the consent form, two questionnaires collecting demographic and health information, and the Geriatric Depression Scale, and they completed the Short Blessed Test. Participants then performed the release-from-PI operation span task, with the processing speed tasks distributed through the test as discussed above. Participants then completed the WASI (first Vocabulary, then Matrix Reasoning). After a short break, participants performed the traditional operation span task followed by the RAPM.

RESULTS

Age Differences in PI and Working Memory

The first set of analyses compares the group of young adults (ages 18–29, n = 67) with the group of older adults (ages 60–79, n = 67) using an extreme groups design. A second set of analyses was then conducted using the older and very old adults (ages 60–95, n = 86) in order to examine age effects in older adulthood using age as a continuous variable. [2]

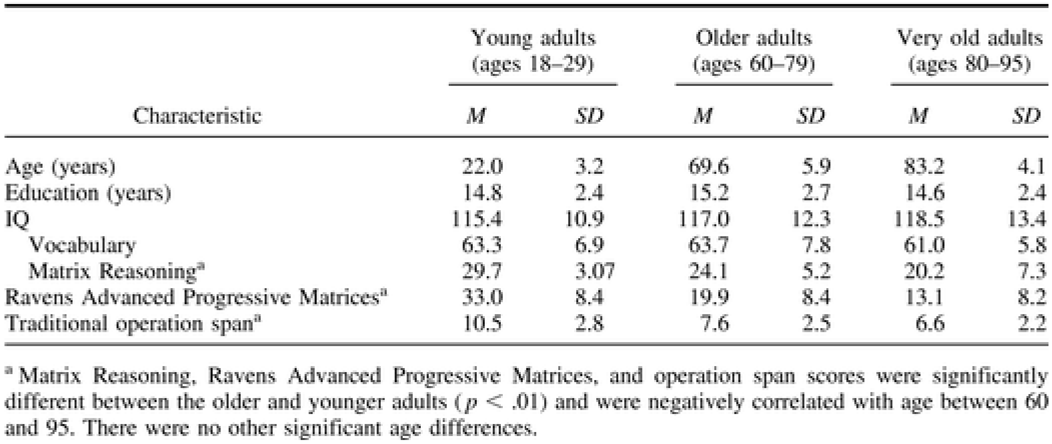
Extreme Groups Analysis

Participant characteristics

Participant characteristics and test scores are presented in Table 1. Older and younger adults did not differ in years of education or age-normed IQ but did show the usual age difference in abstract reasoning (Matrix Reasoning and RAPM raw scores) and traditional operation span scores. It may be noted that the older and younger adults were well matched in terms of IQ, but the IQ means of both older and younger samples were in the high-average range. Although this higher range of ability may limit generalizability, the scores of older adults are not out of line from those in previous cognitive aging research. For example, a meta-analysis by Verhaeghen (2003) showed that WAIS-R vocabulary scores for older adults in previously published research averaged around 60, equivalent to a (WAIS-R) scaled score of 14, or more than one standard deviation above the average for the population.

Table 1

Means and Standard Deviations for Participant Characteristics



	Young adults (ages 18–29)		Older adults (ages 60–79)		Very old adults (ages 80–95)	
Characteristic	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age (years)	22.0	3.2	69.6	5.9	83.2	4.1
Education (years)	14.8	2.4	15.2	2.7	14.6	2.4
IQ	115.4	10.9	117.0	12.3	118.5	13.4
Vocabulary	63.3	6.9	63.7	7.8	61.0	5.8
Matrix Reasoning[a]	29.7	3.07	24.1	5.2	20.2	7.3
Ravens Advanced Progressive Matrices[a]	33.0	8.4	19.9	8.4	13.1	8.2
Traditional operation span[a]	10.5	2.8	7.6	2.5	6.6	2.2

[a] Matrix Reasoning, Ravens Advanced Progressive Matrices, and operation span scores were significantly different between the older and younger adults ($p < .01$) and were negatively correlated with age between 60 and 95. There were no other significant age differences.

Release-from-PI operation span

Figure 1 shows the proportions of the young and older adult groups who answered each trial correctly. To test the hypothesis that older adults would show greater PI susceptibility than younger adults, we conducted a 3 (list length: two, three, or four words) × 4 (buildup cycle: first, second, third, or fourth) × 3 (PI level: low, moderate, or high) × 2 (age group: young vs. old) ANOVA on the release-from-PI operation span data. Because only monotonic decreases in performance across list length and PI level were theoretically interpretable, only main effects and interactions that show significant linear or quadratic components (as determined by significant within-subject linear contrasts) are discussed below.[3] The complete ANOVA results are reported in Table 2.

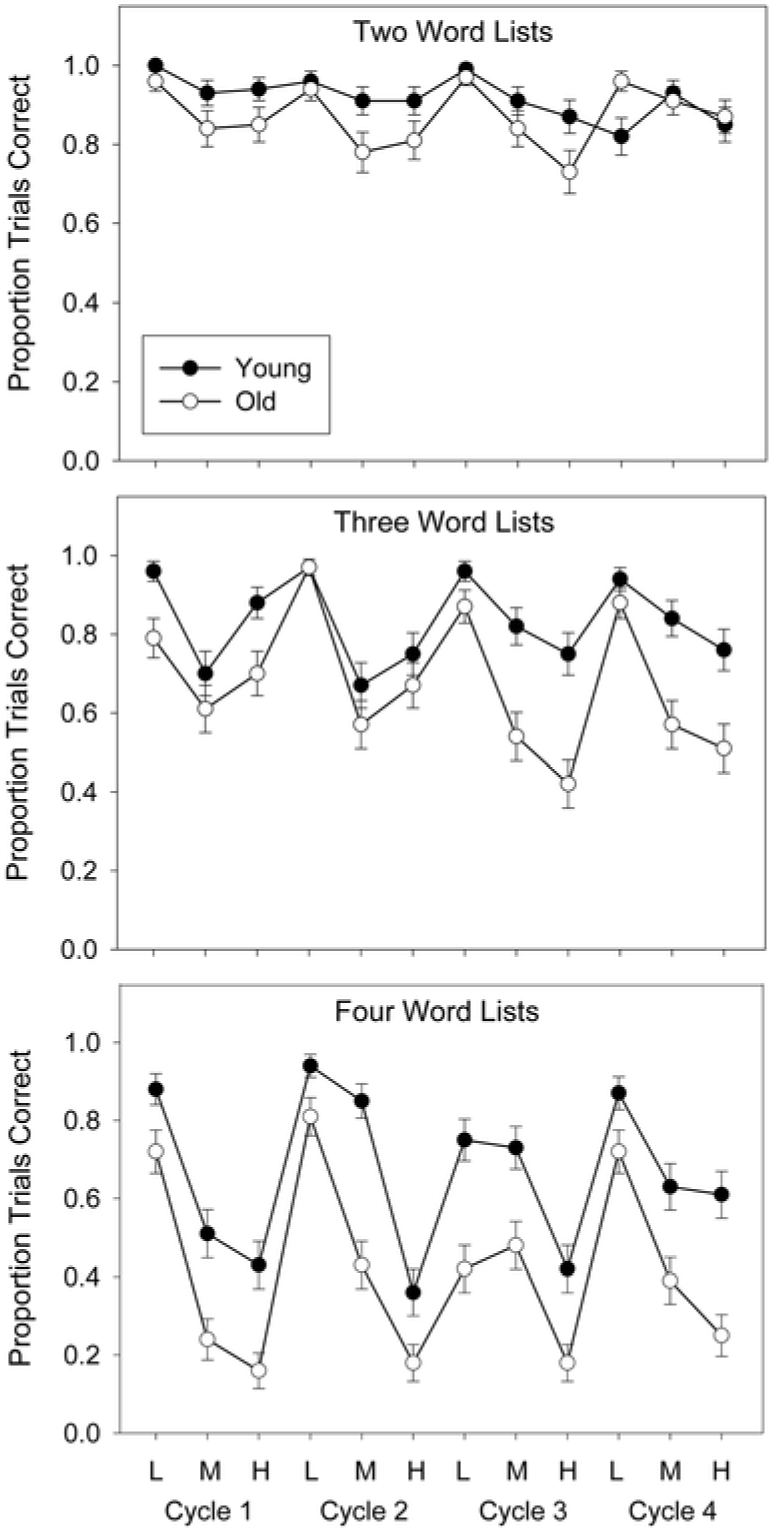


Figure 1. Results from the release–from–proactive interference (PI) operation span. Each panel shows the proportion of participants in each group who answered correctly for all 12 trials at a particular series length. Within the 12 trials, there are four cycles of PI buildup (Cycles 1–4) and

3 trials in each cycle (L = low PI, M = moderate PI, and H = high PI). Error bars are standard errors.

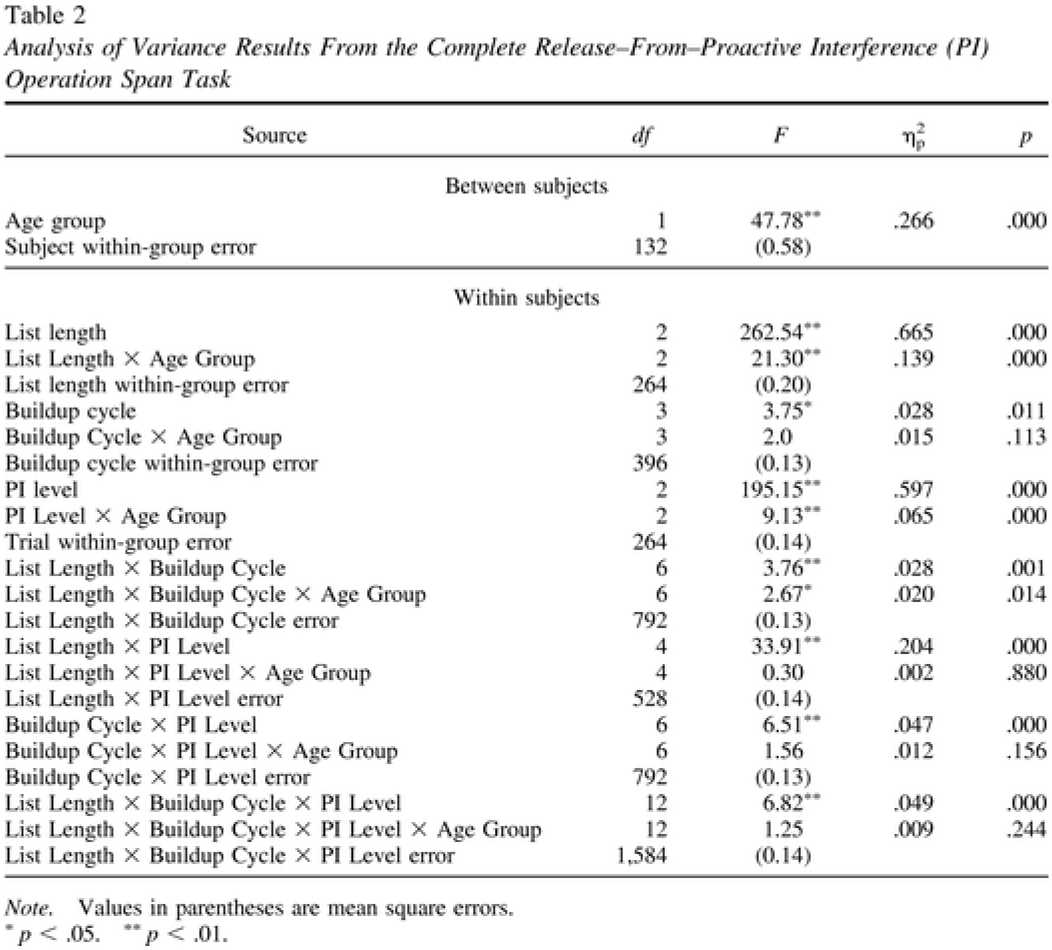


Table 2

Analysis of Variance Results From the Complete Release–From–Proactive Interference (PI) Operation Span Task

Source	*df*	*F*	η_p^2	*p*
Between subjects				
Age group	1	47.78^{**}	.266	.000
Subject within-group error	132	(0.58)		
Within subjects				
List length	2	262.54^{**}	.665	.000
List Length × Age Group	2	21.30^{**}	.139	.000
List length within-group error	264	(0.20)		
Buildup cycle	3	3.75^{*}	.028	.011
Buildup Cycle × Age Group	3	2.0	.015	.113
Buildup cycle within-group error	396	(0.13)		
PI level	2	195.15^{**}	.597	.000
PI Level × Age Group	2	9.13^{**}	.065	.000
Trial within-group error	264	(0.14)		
List Length × Buildup Cycle	6	3.76^{**}	.028	.001
List Length × Buildup Cycle × Age Group	6	2.67^{*}	.020	.014
List Length × Buildup Cycle error	792	(0.13)		
List Length × PI Level	4	33.91^{**}	.204	.000
List Length × PI Level × Age Group	4	0.30	.002	.880
List Length × PI Level error	528	(0.14)		
Buildup Cycle × PI Level	6	6.51^{**}	.047	.000
Buildup Cycle × PI Level × Age Group	6	1.56	.012	.156
Buildup Cycle × PI Level error	792	(0.13)		
List Length × Buildup Cycle × PI Level	12	6.82^{**}	.049	.000
List Length × Buildup Cycle × PI Level × Age Group	12	1.25	.009	.244
List Length × Buildup Cycle × PI Level error	1,584	(0.14)		

Note. Values in parentheses are mean square errors.
$^{*}p < .05.$ $^{**}p < .01.$

Analysis of Variance Results From the Complete Release–From–Proactive Interference (PI) Operation Span Task

The primary interest was in age differences in buildup of PI, represented here by the PI Level × Age Group interaction. As shown in Figure 1 and Table 2, there was a significant PI Level × Age Group interaction, which showed significant linear and quadratic components, $F(1, 132) = 12.87$, $p < .001$, $\eta_p^2 = .09$, and $F(1, 132) = 5.22$, $p < .05$, $\eta_p^2 = .04$, respectively. As shown in Figure 1, both younger and older adults showed a decrease in performance across PI level, with the decrease more pronounced in older adults, and the age difference more pronounced from the low PI to moderate PI trials than from the moderate PI to high PI trials.

It is also of note that list length showed significant linear, $F(1, 132) = 442.59$, $p < .001$, $\eta_p^2 = .77$, and quadratic, $F(1, 132) = 9.08$, $p < .01$, $\eta_p^2 = .06$, trends, with performance decreasing slightly less from two- to three-word lists than from three- to four-word lists. In addition, there was a significant linear List Length × Age Group interaction, $F(1, 132) = 36.32$, $p < .001$, $\eta_p^2 = .22$, indicating that the age difference increased with list length. Finally, there was a significant linear List Length × PI Level interaction, $F(1, 132) = 99.15$, $p < .001$, $\eta_p^2 = .43$, indicating that PI built up faster at longer list lengths. It is important to note that the List Length × PI Level × Age Group interaction was not significant, suggesting that the age difference in PI buildup did not differ significantly at each list length.

Taken together, the present results support the hypothesis that older adults show greater buildup of PI than do young adults. It is of interest, however, that the age difference in working memory performance was not eliminated on the low PI trials, $t(132) = 4.33$, $p < .001$. There are at least two possible interpretations of this result: (a) Older adults did not show as much PI release as young adults, and/or (b) older adults have a smaller basic capacity than do young adults.

To examine the possibility that older adults were not showing PI release to the same extent as are young adults, we conducted an analysis to compare the proportion of the three first trials (i.e., the first trial at each series length) answered correctly to the proportion of the nine release trials answered correctly. A 2 (age group: young vs. old) × 2 (PI level: first vs. release) ANOVA revealed a main effect of age group, $F(1, 132) = 20.64$, $p < .001$, $\eta_p^2 = .14$, but no effect of trial, $F(1, 132) = 0.41$, $p = .52$, $\eta_p^2 = .00$, and no Age Group × PI Level interaction, $F(1, 132) = 2.35$, $p = .13$, $\eta_p^2 = .02$. Consistent with these results, the young adults recalled 94% of the first trials correctly and 91% of the release trials correctly, whereas the older adults remembered 82% of the first trials correctly and 84% of the release trials correctly. Thus, the age difference on low PI trials does not appear to be due to older adults' failure to experience release from PI. Supporting this conclusion are the prior ANOVA results, which indicated no interaction of buildup cycle and age on operation span performance: If older adults were showing lack of release across trials, one might have expected age differences to be larger on Cycle 4 than on Cycle 1.

The possibility that older adults have a smaller basic working memory capacity is discussed further in the Discussion. For now, however, it is important to rule out the possibility that the greater age differences in the higher PI trials is not merely an exaggeration in baseline age differences. To do this, we conducted a hierarchical regression analysis using the high PI trials as the dependent variable, with low PI trial performance and age group entered hierarchically as predictors. If baseline age differences account for the age differences in the high PI trials, the effect of age group on high PI trial performance should be eliminated after controlling for low PI trial performance. This was not the case: Age group accounted for an additional (and significant) 9.4% of the variance in high PI trials after performance on the low PI trials was taken into account ($\Delta R^2 = .094$), $F(1, 131) = 19.91$, $p < .01$.

Age as a Continuous Variable

Participant characteristics and test scores

Participant characteristics of the additional 19 very old adults are given in the third column of Table 1. Within the entire group of older and very old adults (ages 60–95), age was significantly negatively correlated with Matrix Reasoning (raw) scores ($r = -.26$, $p < .05$), RAPM, ($r = -.29$, $p < .01$), and traditional operation span ($r = -.31$, $p < .01$) but not with education, vocabulary, or age-normed IQ.

Release-from-PI operation span results

To further examine the relationship between age, PI buildup, and working memory performance, we conducted an additional analysis on the group of older and very old adults using age as a continuous variable. To simplify the analysis, we collapsed the data over list length and buildup cycle. This analysis was conducted using the repeated measures general linear model (GLM) in SPSS, with PI level (low PI vs. moderate PI vs. high PI) as a (categorical) within-subjects repeated measures variable and (centered) age as a continuous covariate.[4]

As in the extreme groups analysis, we found a significant Age × PI Level interaction, $F(2, 168) = 3.82$, $p < .05$, $\eta_p^2 =.04$. Unlike the previous analysis, the PI Level × Age effect only showed a quadratic trend, $F(1, 84) = 7.15$, $p < .01$, $\eta_p^2 = .08$, not a linear one, $F(1, 84) = 0.07$, $p = .80$, $\eta_p^2 = .00$. To illustrate this interaction, Figure 2 plots performance on the release-from-PI operation span in older adults ages 60–69, 70–79, and 80 and up. Here, it can be seen that PI built up faster in the oldest adults but quickly reached an asymptote, resulting in a larger age difference in the moderate PI trials than in the low or high PI trials. In addition, individuals in their 60s showed much less PI buildup than people in their 70s and up.

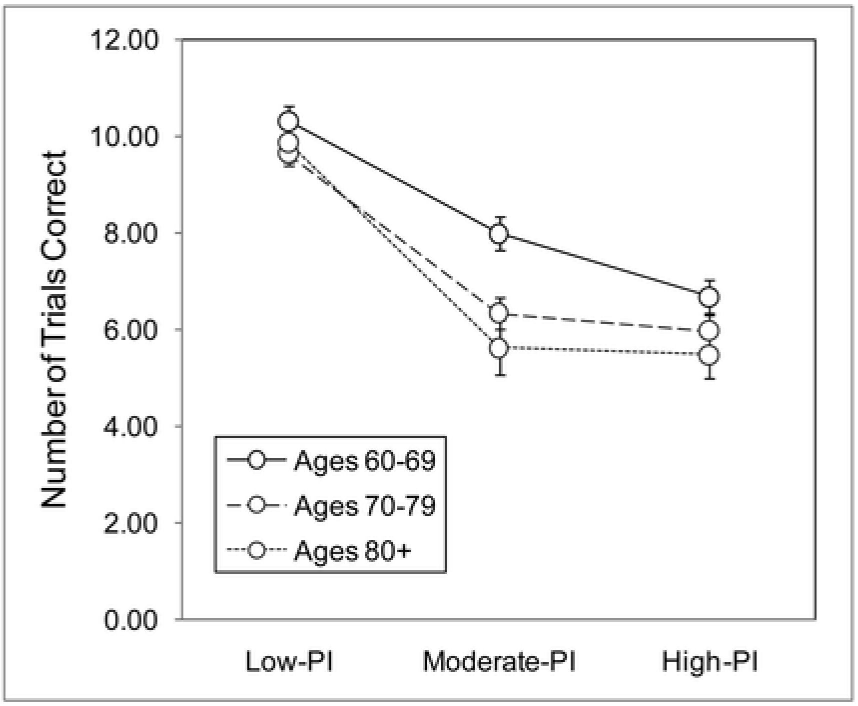


Figure 2. Results from the release–from–proactive interference (PI) operation span within the full older adult group (ages 60–95). Results are shown collapsed over series length and buildup cycle. Error bars are standard errors.

Individual Differences in PI and Working Memory

Individual differences in PI and working memory were first examined separately in the young (ages 18–29, *n* = 67) and older adults (ages 60–79, *n* = 67),[5] following an analysis strategy similar to that used by Bunting (2006). For these analyses, we used the total number of trials correct at each level of PI (low, moderate, and high), collapsing over list length and buildup cycle (e.g., each level of PI was based on performance on 12 total trials). The correlations between reasoning ability and each of the release-from-PI operation span measures were then compared using the Steiger (1980) test for differences between dependent correlations. Hierarchical regression was then used to test whether each increase in PI in the release-from-PI operation span task predicted unique variance in reasoning ability scores. These initial analyses, however, suggested age differences in the relationship between reasoning ability and the buildup of PI. This interaction is directly tested in the final section below.

Inspection of the RAPM and Matrix Reasoning data indicated a strong correlation between the two tests (*r* = .76). Therefore, these two scores were combined into a single abstract reasoning measure by converting each set of raw scores into *z* scores (on the basis of the entire sample), then averaging the two *z* scores for each person. These mean *z* scores were then used in all analyses reported below.

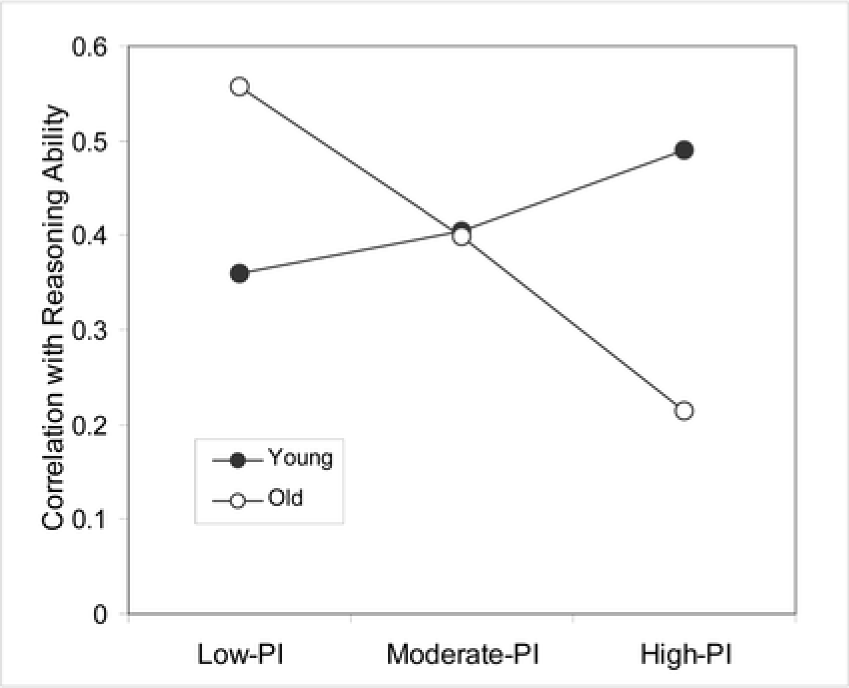


Figure 3. Effect of proactive interference (PI) level on the correlation between working memory and reasoning ability in younger and older adults.

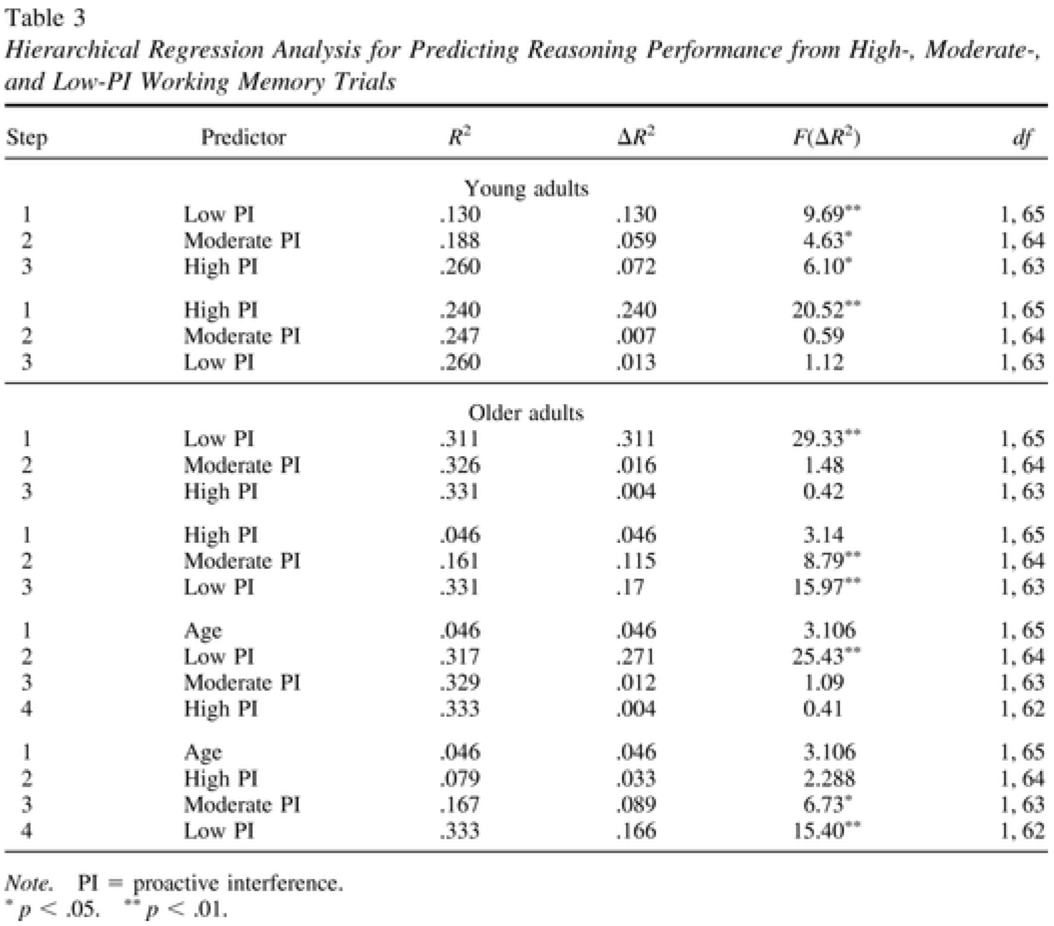


Table 3

Hierarchical Regression Analysis for Predicting Reasoning Performance from High-, Moderate-, and Low-PI Working Memory Trials

Step	Predictor	R^2	ΔR^2	$F(\Delta R^2)$	*df*
		Young adults			
1	Low PI	.130	.130	9.69**	1, 65
2	Moderate PI	.188	.059	4.63*	1, 64
3	High PI	.260	.072	6.10*	1, 63
1	High PI	.240	.240	20.52**	1, 65
2	Moderate PI	.247	.007	0.59	1, 64
3	Low PI	.260	.013	1.12	1, 63
		Older adults			
1	Low PI	.311	.311	29.33**	1, 65
2	Moderate PI	.326	.016	1.48	1, 64
3	High PI	.331	.004	0.42	1, 63
1	High PI	.046	.046	3.14	1, 65
2	Moderate PI	.161	.115	8.79**	1, 64
3	Low PI	.331	.17	15.97**	1, 63
1	Age	.046	.046	3.106	1, 65
2	Low PI	.317	.271	25.43**	1, 64
3	Moderate PI	.329	.012	1.09	1, 63
4	High PI	.333	.004	0.41	1, 62
1	Age	.046	.046	3.106	1, 65
2	High PI	.079	.033	2.288	1, 64
3	Moderate PI	.167	.089	6.73*	1, 63
4	Low PI	.333	.166	15.40**	1, 62

Note. PI = proactive interference.
$^{*} p < .05$. $^{**} p < .01$.

Hierarchical Regression Analysis for Predicting Reasoning Performance from High-, Moderate-, and Low-PI Working Memory Trials

The one caveat to these results is that despite the high reliability (α = .80) of the release-from-PI operation span task for the younger adults in the present study, the reliability of the low PI trials, computed separately, was relatively low (α = .32), and performance on these trials was near ceiling (92% of trials correct). These factors could have resulted in an underestimation of the correlation between reasoning ability and performance on low PI trials. It should be noted, however, that the present findings replicate those of Bunting (2006), who used a similar release-from-PI operation span procedure and also found that performance on the low PI trials was only weakly correlated with reasoning (*r* = .24), despite the fact that the reliability for low PI trials in the Bunting young adult study was much higher (α = .63) and performance much lower. Moreover, the reliability of the moderate and high PI trials in the present study were relatively high (αs of .65 and .60, respectively), and performance was not near ceiling (78% correct and 71% correct, respectively). Thus, the fact that the high PI trials accounted for additional variance in reasoning ability not accounted for by the two lower PI levels (which contributed no variance not accounted for by the high PI trials) clearly supports the hypothesis that working memory performance under high PI conditions predicts reasoning ability better than performance when there is less PI.

Older Adults

To take into account the larger age range represented in the older adults, we performed analyses both with and without age as a covariate. In the older adults, the correlation between the release-from-PI operation span and reasoning ability decreased as the amount of PI increased, both for zero-order correlations (see Figure 3) and for the partial correlations with the effect of age removed (*r*s = .54, .35, and .19 for low, moderate, and high PI trials, respectively). In fact, the correlation between the high PI trials and reasoning ability did not reach significance in the older adults (*p* = .08). This is in contrast to the young adult data, in which performance on the high PI trials was most predictive of reasoning ability. In the older adults, the correlation between the low PI trials and reasoning ability was significantly higher than the correlation between the high PI trials and reasoning ability (*z* = 2.91, *p* < .01, two-tailed), but none of the other differences reached significance (*z* = 1.56 for low PI vs. moderate PI; *z* = 1.45 for moderate PI vs. high PI).

Perhaps surprisingly, hierarchical regression analyses of the older adult data revealed that neither moderate nor high PI trials accounted for any variance in reasoning ability after performance on the low PI trials was controlled. Moreover, the low PI trials did account for additional variance in reasoning ability that was not accounted for by the moderate or high PI trials (see the bottom of Table 3). Finally, the correlations between reasoning ability and (a) the total release-from-PI operation span score (*r* = .47) and (b) the traditional operation span score (*r* = .40) are lower than the correlation between the reasoning ability and the low PI trials alone (*r* = .56). The fact that the low PI trials alone predict reasoning ability slightly better than the traditional operation span suggests that high levels of PI may simply add noise to the predictive ability of operation span in older adults, at least when the dependent variable is inductive reasoning ability.

As was the case with the younger adults, the overall reliability of the release-from-PI operation span was relatively high (α = .77), but the reliability was lower for the different PI levels considered separately. In contrast to the younger adult data, however, the different PI levels showed relatively similar reliability (α = .49 for low PI trials, α = .52 for moderate PI trials, and α = .56 for high PI trials). Thus, the pattern of results depicted in Figure 3, like the hierarchical regression results for older adults, is clearly inconsistent with the hypothesis that working memory performance under high PI conditions is the best predictor of older adults' reasoning ability. The low correlation between the low PI trials and reasoning ability also does not appear to be the result of a floor effect in low PI trial performance because participants averaged 53% correct on these trials, considerably far from floor.

Interaction Analysis

The results of the above analyses suggested an unexpected interaction between age group, reasoning ability, and PI buildup. This impression was confirmed with a PI Level × Age Group × Reasoning Ability GLM analysis on the release-from-PI operation span data in which reasoning ability was entered as a continuous covariate. This analysis verified the presence of a significant Age Group × Reasoning Ability × PI Level interaction, $F(2, 260) = 5.79$, $p < .01$, $\eta_p^2 = .04$.[6] Separate GLM analyses for the two age groups confirmed that in the young adult group, there was a significant Reasoning Ability × PI Level interaction, $F(2, 130) = 4.78$, $p < .01$, $\eta_p^2 = .07$, but there was no such interaction in the older adult group, $F(2, 130) = 2.11$, $p = .13$, $\eta_p^2 = .03$. These results are consistent with the preceding regression analyses in showing that PI buildup is related to reasoning ability in the young adults but not in the older adults.

The Age Group × Reasoning Ability × PI Level interaction is depicted in Figure 4, which plots the release-from-PI operation span data as a function of age group and reasoning ability. For illustration purposes, the participants were divided into higher and lower reasoning ability groups on the basis of mean *z* scores computed on all of the participants, regardless of age. Participants with higher reasoning ability were those with mean *z* scores greater than 0, and participants with lower reasoning ability were those with mean *z* scores less than 0. As may be seen in Figure 3, older adults with higher reasoning ability ($n = 15$) performed similarly to the younger adults with higher reasoning ability ($n = 54$) on low PI trials but showed a greater drop in performance as PI built up. In contrast, the younger adults with lower reasoning ability ($n = 14$) performed better on the low PI trials than the older adults with lower reasoning ability ($n = 52$) did but showed an equivalent drop in performance as PI increased. Thus, the size of the age difference in buildup of PI clearly depends on the level of reasoning ability in the younger and older adults.

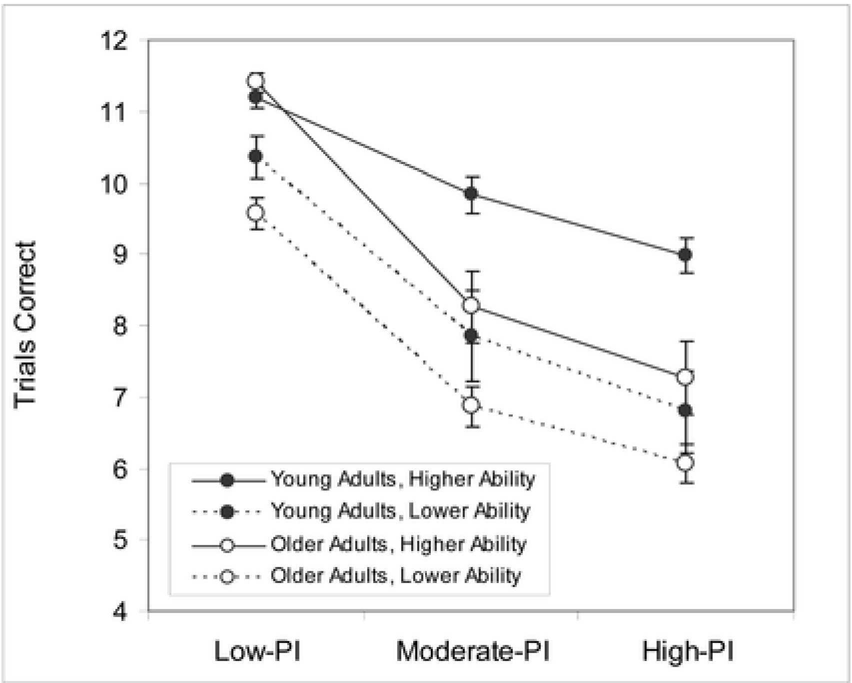


Figure 4. Effect of proactive interference (PI) level on working memory performance as a function of age group and reasoning ability. Error bars are standard errors.

DISCUSSION

The current study was designed to answer two questions: (a) Do age differences in the ability to resist PI contribute to age differences in working memory, and (b) does working memory performance under high PI conditions predict individual differences in abstract reasoning ability better than performance under low PI conditions in both older and younger adults? With respect to the first question, the results are relatively clear-cut and supportive of theories that highlight age differences in the susceptibility to interference. With respect to the second question, however, the results are somewhat more problematic and suggest some limitations to inhibitory theories.

Age Differences in PI

Comparisons of younger and older adults' performance on the release-from-PI operation span task revealed that PI built up more quickly in the older group, as indicated by the significant Age Group × PI Level interactions in the ANOVA. Hierarchical regression analyses showed that although age differences were not completely absent on low PI trials, an age difference in the high PI trials remained even after controlling for low PI trial performance. This suggests that resistance to PI contributes to age differences in working memory ability over and above possible age differences in basic memory capacity. Similar results were obtained when age was analyzed as a continuous variable within the older adults. There was a significant Age × PI Level interaction, indicating that the rate of PI buildup increased as age increased. The age difference was larger on the moderate PI trials than on either the low or high PI trials, however, suggesting that as older adults age, the buildup of PI across trials becomes even more rapid but then levels off.

It may be noted that the young adults' performance on low PI trials was close to ceiling (92% correct), and this could have contributed to the Age Group × PI Level ANOVA interaction. Ceiling effects were not an issue, however, in the GLM analysis that examined the working memory performance of both the older and very old adults (60–95 years of age) and in which age was treated as a continuous variable. The performance of the older adults on low PI trials was generally well below ceiling (82% correct) regardless of their age, and yet an Age × PI Level interaction was still observed. Thus, the strongest evidence for the hypothesis of age differences in the buildup of PI comes from analysis of the data from all of the older adults, although the results of the comparison of young and older adults are also consistent with this hypothesis.

Overall, the findings of the current study support the view that age differences on working memory tasks are influenced by age differences in the ability to resist PI. It is worth noting, however, that the age difference in PI by itself cannot completely explain the results from the release-from-PI operation span task. Differences between the performance of older and younger adults were found on the low PI trials, even on the very first trial at each series length when no PI from previous trials could have been present. In addition, the Age Group × List Length interaction was significant, whereas the Age Group × List Length × PI Level interaction was not, indicating that the age difference increased from shorter to longer lists, regardless of the amount of PI that was present. Taken together, these results suggest that in addition to having more difficulty with PI, older adults also have a smaller basic working memory capacity.

The current results, specifically the presence of age differences even on low PI trials, may be contrasted with the results of ascending versus descending reading span comparisons in previous studies, which showed no age differences with descending presentation. One possible reason for this discrepancy is that presenting items in ascending order may be particularly problematic for older adults for reasons over and above their greater susceptibility to PI. For example, the progressive increases in difficulty created by increasing list lengths may create more anxiety in older adults than in younger adults.

An alternative explanation is that although older adults benefit from manipulations that directly reduce PI (e.g., changes in material or changes in task), they may also incur a small cost to their performance as a result of difficulties in switching from one task or set to another (e.g., Kray & Lindenberger, 2000). For example, May et al. (1999) found that although older adults benefited from both breaks between trials and descending presentation separately, combining the two did not further improve their spans and, in fact, actually increased the age difference in performance. Similarly, the current study combined switches in material (categories of words) and switches in task (performance of an unrelated task between series lengths) to reduce PI, but age differences were still present on low PI trials after a change in word category, and even on the initial trial at each series length following a break—the very trials that would be most affected by difficulties in set switching. One way for future studies to address these possibilities would be to repeat the current procedure with list lengths presented in descending order. Determining whether age differences in the low PI trials still persisted under these conditions would shed further light on the source of older adults' problems with ascending series on working memory tasks.

Individual Differences in PI and Abstract Reasoning

The picture of the relationship between susceptibility to PI and individual differences in reasoning ability that emerges from the results of the current study is somewhat complicated and suggests that a refinement to current views of the role of PI in higher order cognition is needed. Although the young adult data were consistent with previous findings, the older adults showed a markedly different pattern. In the young adults, abstract reasoning was best predicted by performance on the high PI trials of the release-from-PI operation span task, replicating Bunting (2006). For the older adults in the current study, however, the high PI trials were not even a significant predictor of reasoning ability. Indeed, the best predictor of reasoning ability in the older adults was their performance on low PI trials. Moreover, the moderate PI trials were not significant predictors of reasoning once low PI trial performance was controlled. The different patterns of correlations between working memory performance and reasoning observed in the older and younger adults was confirmed by a three-way interaction between PI level, age, and reasoning ability. It should be noted that because of the homogeneity in susceptibility to PI among older adults of different reasoning ability, in contrast to the heterogeneity observed among young adults (see Figure 4), other factors necessarily emerge as predictors of older adults' reasoning ability.

The present results may be contrasted with those of Lustig et al. (2001), who observed that older adults' reading span performance under high PI conditions (ascending series) was a better predictor of their prose recall than performance under low PI conditions. It is possible that the current findings are specific to predicting reasoning ability because of the set switching embedded in reasoning tests like RAPM and Matrix Reasoning. Indeed, previous research suggests that task switching is strongly related to abstract reasoning and less related to general knowledge or memory (Kray & Lindenberger, 2000). Both RAPM and Matrix Reasoning present a series of problems that differ in the rules needed to solve them. This requires an ability to flexibly switch between mental sets both within and between problems as one considers potential solutions. Similarly, the low PI trials in the release-from-PI operation span task require abandoning the previous category of words and determining what the new category is.

Thus, it may be that what best distinguishes among older adults of different levels of reasoning ability is the ability to abandon one mental set in favor of another, Of course, this switching ability could also be important in young adults. As previously mentioned, the low PI trials in the current study were predictive of reasoning ability in young adults. Alternatively, it may be simply that the relationship between PI and higher level cognition in older adults is relatively domain specific. In the Lustig et al. (2001) study, for example, both the PI task (reading span) and the higher level task (prose recall) were verbal memory tasks, whereas in the current study, we used a verbal PI task (operation span) to predict a more spatial reasoning task (matrix reasoning). In the past, our laboratory has found that spatial tasks predict reasoning ability in older adults better than do verbal tasks (Hale et al., 2007), and it may be that a high PI task involving spatial memory material may be a better predictor of the kind of reasoning ability tested in the current study.

Implications and Conclusions

The current study has important implications for both researchers and clinicians interested in cognitive aging. First, the current results suggest that the use of working memory procedures that produce considerable PI may not be the best option when the goal is to predict reasoning ability in older adults. Instead, procedures that involve relatively little PI appear to best distinguish among older adults of different levels of reasoning ability in older adults. In addition, the current results highlight the fact that in cross-sectional comparisons, whether one finds an age-related deficit specific to a particular ability (e.g., resistance to PI) may depend on both the characteristics of the older adult group being tested and the characteristics of the young adult group being tested. For example, if the current study had only sampled younger and older adults with higher reasoning ability, the results would have led to the conclusion that there are age differences in PI buildup. If the current study had contained only young and older adults with average reasoning ability, however, the results would have led to the conclusion that there are no age differences in PI buildup! Taken together, the present results clearly demonstrate how age and abstract reasoning ability interact in determining the effects of PI on working memory performance. Finally, the present findings have implications for the relations between age, PI, working memory, and reasoning. Specifically, the finding that older adults' reasoning was better predicted by their working memory performance under low PI conditions than by their performance under high PI conditions raises questions regarding the general role of susceptibility to PI in individual differences in higher cognitive function among older adults.

NOTES

1. It may be noted that both the Battig and Montague (1969) norms and the Balota et al. (2002) naming and lexical decision times were based on groups of younger adults only. The Battig and Montague norms were collected in the 1960s, however, when some members of our older adult group would have been in their 20s, perhaps making the norms more relevant for them than our current young adult group.
2. The very old adults were not included in the young versus old comparison in order to have equal numbers of participants in each group (young and old). Including the very old group in the young versus old comparison, however, does not significantly change any of the results.
3. As may be seen in the full ANOVA table, there were several higher order interactions involving buildup cycle. However, there was no significant linear or quadratic component

to the buildup cycle effect, and the higher order interactions showed no theoretically interpretable pattern. These interactions were likely due to differences in memory and PI buildup for different categories of words. With respect to the significant List Length × Buildup Cycle × Age Group interaction, this appears to be due to age differences in memory for different categories of words. For example, older adults outperform young adults on the fourth buildup cycle of the two-word lists (vegetables), and smaller age differences are seen in the first and second buildup trials of the three-word lists (musical instruments and kitchen appliances) than other categories. Important for the present study, however, buildup cycle does not interact with PI level and age, indicating that collapsing across buildup cycles should not significantly influence subsequent analyses.

4. Treating age as a continuous covariate provides more power to detect the Age × PI Level interaction than does dividing age into groups and analyzing the data using ANOVA. This approach has been used similarly in previous research to detect interactions between continuous and categorical variables (e.g., Tamir, 2005). We note that in the current study, an analysis in which age was treated was categorical (by decade, as depicted in Figure 2) also yielded a significant Age Group × PI Level interaction, $F(4, 166) = 2.60$, $p = .04$.
5. Including the very old adults in the analysis did not significantly change the results.
6. As in the previous GLM analysis, dichotomizing the reasoning variable as depicted in Figure 4 also yielded a significant Age Group × Reasoning Ability × PI Level interaction, $F(2, 260) = 3.64$, $p = .03$.

REFERENCES

Baddeley, A. (1986). *Working memory*. New York: Oxford University Press.

Balota, D. A., Cortese, M. J., Hutchinson, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2002). *The English lexicon project: A Web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. Retrieved June 7, 2004, from Washington University Web site: http://elexicon.wustl.edu

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*, 1–46.

Bowles, R. P., & Salthouse, T. A. (2003). Assessing the age-related effects of proactive interference on working memory tasks using the Rasch model. *Psychology and Aging*, *18*, 608–615.

Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 183–196.

Daneman, M., & Carpenter, P. M. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422–433.

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

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51*, 136–158.

Hale, S., Myerson, J., Emery, L., Lawrence, B. M., & DuFault, C. (2007). Variation in working memory across the lifespan. In A. R. A.Conway, C.Jarrold, M. J.Kane, A.Miyake, & N.Towse (Eds.), *Variation in working memory* (pp. 194–224). New York: Oxford University Press.

Hedden, T., & Yoon, C. (2006). Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology*, *20*, 511–528.

Hertzog, C., Dixon, R. A., Hultsch, D. F., & MacDonald, S. W. S. (2003). Latent change models of adult cognition: Are changes in processing speed and working memory associated with changes in episodic memory?*Psychology and Aging*, *18*, 755–769.

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

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging*, *15*, 126–147.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*, 199–207.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory and Cognition*, *27*, 759–767.

Myerson, J., Emery, L., White, D. A., & Hale, S. (2003). Effects of age, domain, and processing demands on memory span: Evidence for differential decline. *Aging, Neuropsychology, and Cognition*, *10*, 20–27.

Neupert, S. D., Soederberg, L. M., & Lachman, M. E. (2006). Physiological reactivity to cognitive stressors: Variations by age and socioeconomic status. *International Journal of Aging and Human Development*, *62*, 221–235.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., & Smith, A. D. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*, 299–320.

Psychological Corporation. (1997). *WAIS-III administration and scoring manual*. San Antonio, TX: Harcourt Brace.

Psychological Corporation. (1999). *WASI: Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Author.

Raven, J. C., Raven, J., & Court, J. H. (1993). *Manual for Raven's progressive matrices and vocabulary scales, Section 4: Advanced progressive matrices* (1994 ed.). Oxford, England: Oxford Psychologists Press.

Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology*, *84*, 171–199.

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, *8*, 535–543.

Schaie, K. W. (2005). *Developmental influences on adult intelligence: The Seattle longitudinal study*. New York: Oxford University Press.

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric depression scale (GDS): Recent evidence and development of a shorter version. In R. L.Brink (Ed.), *Clinical gerontology: A guide to assessment and intervention* (pp. 165–173). New York: The Hawthorne Press.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.

Tamir, M. (2005). Don't worry, be happy? Neuroticism, trait-consistent affect regulation, and performance. *Journal of Personality and Social Psychology*, *89*, 449–461.

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

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, *54*, 68–80.

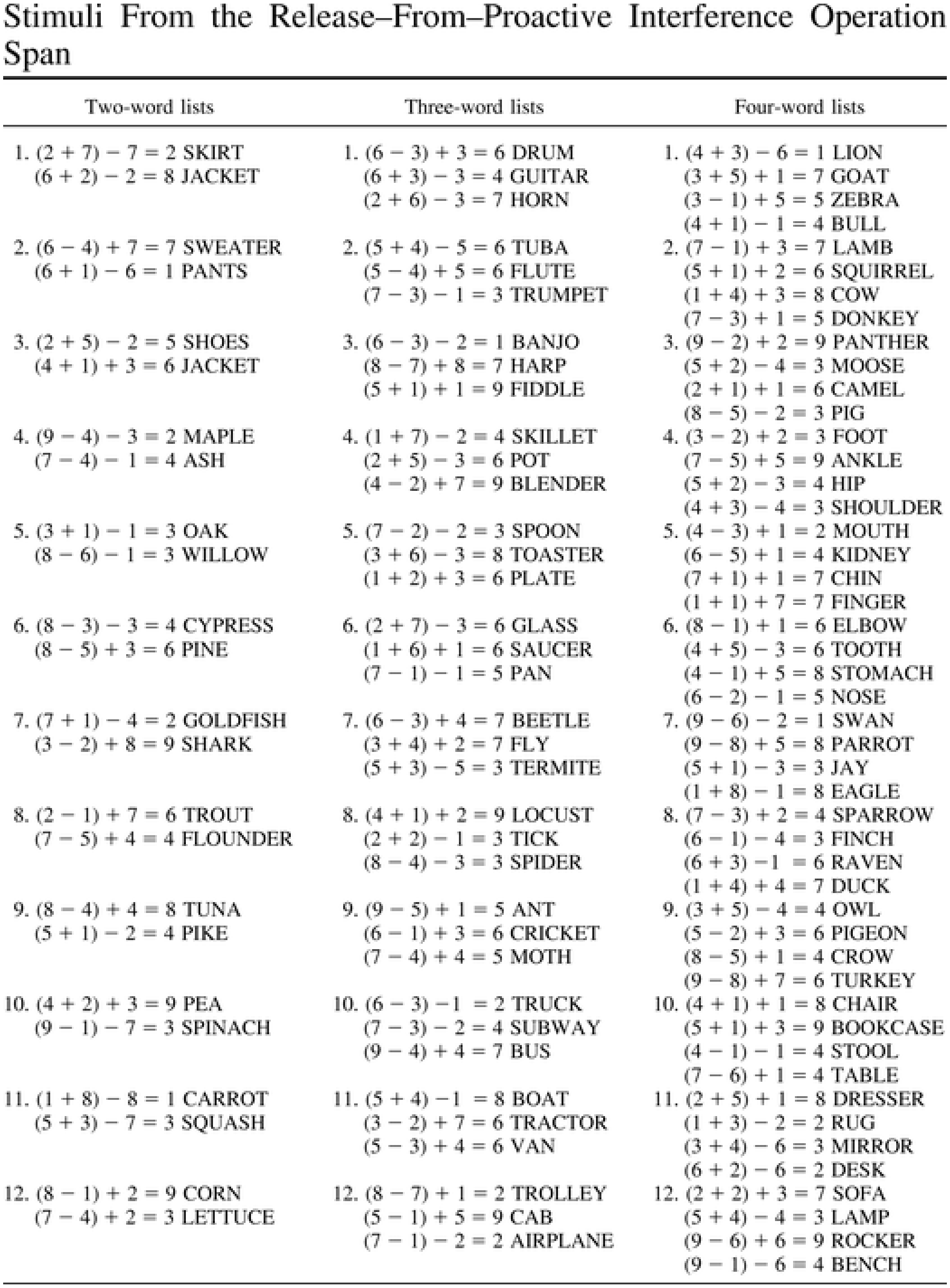
Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, *18*, 332–339.

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

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, *77*, 1–15.

APPENDIX

APPENDIX A: Stimuli From the Release–From–Proactive Interference Operation Span



Stimuli From the Release–From–Proactive Interference Operation Span

Two-word lists	Three-word lists	Four-word lists
1. (2 + 7) − 7 = 2 SKIRT (6 + 2) − 2 = 8 JACKET	1. (6 − 3) + 3 = 6 DRUM (6 + 3) − 3 = 4 GUITAR (2 + 6) − 3 = 7 HORN	1. (4 + 3) − 6 = 1 LION (3 + 5) + 1 = 7 GOAT (3 − 1) + 5 = 5 ZEBRA (4 + 1) − 1 = 4 BULL
2. (6 − 4) + 7 = 7 SWEATER (6 + 1) − 6 = 1 PANTS	2. (5 + 4) − 5 = 6 TUBA (5 − 4) + 5 = 6 FLUTE (7 − 3) − 1 = 3 TRUMPET	2. (7 − 1) + 3 = 7 LAMB (5 + 1) + 2 = 6 SQUIRREL (1 + 4) + 3 = 8 COW (7 − 3) + 1 = 5 DONKEY
3. (2 + 5) − 2 = 5 SHOES (4 + 1) + 3 = 6 JACKET	3. (6 − 3) − 2 = 1 BANJO (8 − 7) + 8 = 7 HARP (5 + 1) + 1 = 9 FIDDLE	3. (9 − 2) + 2 = 9 PANTHER (5 + 2) − 4 = 3 MOOSE (2 + 1) + 1 = 6 CAMEL (8 − 5) − 2 = 3 PIG
4. (9 − 4) − 3 = 2 MAPLE (7 − 4) − 1 = 4 ASH	4. (1 + 7) − 2 = 4 SKILLET (2 + 5) − 3 = 6 POT (4 − 2) + 7 = 9 BLENDER	4. (3 − 2) + 2 = 3 FOOT (7 − 5) + 5 = 9 ANKLE (5 + 2) − 3 = 4 HIP (4 + 3) − 4 = 3 SHOULDER
5. (3 + 1) − 1 = 3 OAK (8 − 6) − 1 = 3 WILLOW	5. (7 − 2) − 2 = 3 SPOON (3 + 6) − 3 = 8 TOASTER (1 + 2) + 3 = 6 PLATE	5. (4 − 3) + 1 = 2 MOUTH (6 − 5) + 1 = 4 KIDNEY (7 + 1) + 1 = 7 CHIN (1 + 1) + 7 = 7 FINGER
6. (8 − 3) − 3 = 4 CYPRESS (8 − 5) + 3 = 6 PINE	6. (2 + 7) − 3 = 6 GLASS (1 + 6) + 1 = 6 SAUCER (7 − 1) − 1 = 5 PAN	6. (8 − 1) + 1 = 6 ELBOW (4 + 5) − 3 = 6 TOOTH (4 − 1) + 5 = 8 STOMACH (6 − 2) − 1 = 5 NOSE
7. (7 + 1) − 4 = 2 GOLDFISH (3 − 2) + 8 = 9 SHARK	7. (6 − 3) + 4 = 7 BEETLE (3 + 4) + 2 = 7 FLY (5 + 3) − 5 = 3 TERMITE	7. (9 − 6) − 2 = 1 SWAN (9 − 8) + 5 = 8 PARROT (5 + 1) − 3 = 3 JAY (1 + 8) − 1 = 8 EAGLE
8. (2 − 1) + 7 = 6 TROUT (7 − 5) + 4 = 4 FLOUNDER	8. (4 + 1) + 2 = 9 LOCUST (2 + 2) − 1 = 3 TICK (8 − 4) − 3 = 3 SPIDER	8. (7 − 3) + 2 = 4 SPARROW (6 − 1) − 4 = 3 FINCH (6 + 3) −1 = 6 RAVEN (1 + 4) + 4 = 7 DUCK
9. (8 − 4) + 4 = 8 TUNA (5 + 1) − 2 = 4 PIKE	9. (9 − 5) + 1 = 5 ANT (6 − 1) + 3 = 6 CRICKET (7 − 4) + 4 = 5 MOTH	9. (3 + 5) − 4 = 4 OWL (5 − 2) + 3 = 6 PIGEON (8 − 5) + 1 = 4 CROW (9 − 8) + 7 = 6 TURKEY
10. (4 + 2) + 3 = 9 PEA (9 − 1) − 7 = 3 SPINACH	10. (6 − 3) −1 = 2 TRUCK (7 − 3) − 2 = 4 SUBWAY (9 − 4) + 4 = 7 BUS	10. (4 + 1) + 1 = 8 CHAIR (5 + 1) + 3 = 9 BOOKCASE (4 − 1) − 1 = 4 STOOL (7 − 6) + 1 = 4 TABLE
11. (1 + 8) − 8 = 1 CARROT (5 + 3) − 7 = 3 SQUASH	11. (5 + 4) −1 = 8 BOAT (3 − 2) + 7 = 6 TRACTOR (5 − 3) + 4 = 6 VAN	11. (2 + 5) + 1 = 8 DRESSER (1 + 3) − 2 = 2 RUG (3 + 4) − 6 = 3 MIRROR (6 + 2) − 6 = 2 DESK
12. (8 − 1) + 2 = 9 CORN (7 − 4) + 2 = 3 LETTUCE	12. (8 − 7) + 1 = 2 TROLLEY (5 − 1) + 5 = 9 CAB (7 − 1) − 2 = 2 AIRPLANE	12. (2 + 2) + 3 = 7 SOFA (5 + 4) − 4 = 3 LAMP (9 − 6) + 6 = 9 ROCKER (9 − 1) − 6 = 4 BENCH