Published in final edited form as:
Mem Cognit. 2004 September ; 32(6): 956–964.

Using immediate memory span to measure implicit learning

Jeffrey D. Karpicke and
Washington University, St. Louis, Missouri

David B. Pisoni
Indiana University, Bloomington, Indiana

Abstract

To avoid some conceptual and methodological pitfalls found in traditional artificial grammar learning tasks, we developed a new method of measuring implicit learning using immediate memory span. Subjects were presented with sequences generated by an artificial grammar and were asked to reproduce the patterns by pressing buttons on a response box. After exposure to these sequences, subjects showed selective improvement in immediate memory span for novel sequences governed by the same grammar. Individual differences in implicit learning covaried with measures of auditory digit span. Subjects with greater immediate memory processing capacity were better able to learn and subsequently exploit the information available in grammatical sequences. Our results are consistent with a detailed episodic coding framework in which implicit learning occurs as an incidental by-product of explicit task performance. Although subjects encode highly detailed information about specific instances, they use different aspects of this information to accomplish different task-specific demands.

While a considerable amount of attention in the implicit learning literature has been devoted to a variety of issues (Stadler & Frensch, 1998), very little research has explored the nature of individual differences in implicit learning. Reber's (1993) influential theory of implicit learning argued that implicit cognitive processes are evolutionarily older than explicit processes and, therefore, implicit learning should show no individual differences and should display little agreement with measures of intelligence. Only a few studies have specifically investigated individual differences in implicit learning, and the findings from these studies support the proposal that implicit cognitive processes are less susceptible to individual differences than explicit cognitive processes. However, failures to find systematic variations in performance among individuals in implicit learning tasks can be attributed to conceptual and methodological problems. Specifically, measures of classification performance in artificial grammar learning tasks may not be very sensitive to individual differences. In addition, with the exception of some recent studies (e.g., Destrebecqz &Cleeremans, 2001), most research on implicit learning has assumed that behavioral tasks are "process-pure"—that a given task exclusively involves either implicit or explicit knowledge. In this article, we describe a new method for separating implicit and explicit learning processes, largely on the basis of process dissociation logic (Jacoby, 1991), by measuring implicit learning effects in an immediate memory span task.

In the traditional artificial grammar learning task, subjects are asked to memorize letter strings generated by an artificial grammar. After this learning phase, subjects are informed

Correspondence should be addressed either to J. D. Karpicke, Department of Psychology, Washington University, Campus Box 1125, St. Louis, MO 63130, or to D. B. Pisoni, Department of Psychology, Indiana University, 1101 E. 10th Street, Bloomington, IN 47405-7007 (karpicke@wustl.edu or pisoni@indiana.edu).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

that the letter strings they have just learned were generated according to a complex system of grammatical rules, although no information is given about the nature of the rules. Subjects are then presented with novel letter strings and required to decide whether or not each string follows the grammatical rules, without receiving any feedback about their decisions. Half of the letter strings are new grammatical strings that were not presented during the learning phase, and half are randomly generated nongrammatical strings. The main finding, replicated several times in the literature (Reber, 1993), is that subjects can classify novel letter strings as grammatical or nongrammatical well above chance without displaying conscious verbalizable knowledge about the rules of the grammar when questioned following the experiment.

Very few studies have investigated individual differences and variability in implicit learning using the artificial grammar learning task. Reber, Walkenfeld, and Hernstadt (1991) compared performance on the traditional artificial grammar learning task with performance on an explicit problem-solving task and also with subtests from the Wechsler Adult Intelligence Scale–Revised (WAIS–R). They found a significant positive correlation between performance on the explicit problem-solving task and IQ, but no significant correlation between the artificial grammar learning task and IQ or between the implicit and explicit tasks.

Mayberry, Taylor, and O’Brien-Malone (1995) also found that performance among school-age children on an implicit learning task was not related to IQ, but that performance on an explicit task was. McGeorge, Crawford, and Kelly (1997) tested subjects from age 18 to 77 and found, once again, that performance on the implicit learning task did not correlate with IQ, while performance on an explicit test did. McGeorge et al. also found that while there were no differences in performance on the implicit test with increasing age, performance on the explicit test decreased with age. All of these findings provide support for the argument that implicit learning does not covary with either explicit learning or traditional psychometric measures of intelligence.

Nearly all of the discussions concerning individual differences in implicit learning conclude that they do not exist. This follows from the theoretical arguments proposed by Reber (1993, 1997). However, the empirical studies described above have demonstrated only that implicit measures of learning do not covary with explicit measures of learning, which is not evidence for the nonexistence of individual differences in implicit learning processes. Furthermore, a relationship between explicit tests and IQ, and none between implicit tests and IQ, is a foregone conclusion, since intelligence tests are specifically designed to measure the same construct thought to underlie explicit tests but not to measure implicit cognitive processes. None of the studies by Reber et al. (1991), Mayberry et al. (1995), or McGeorge et al. (1997) was designed to reveal or explain variability in implicit learning processes. Measures of grammaticality judgments in artificial grammar learning may simply be insensitive to variability among individuals.

Recently, Reber and Allen (2000) suggested that, “If we are to come to some conclusions about interindividual differences in implicit learning, we need to examine individual performances on implicit and explicit tasks which use a common metric” (p. 241). Their argument is for a task dissociation: A common metric is needed so that a direct comparison can be made between implicit and explicit tasks. However, this is the same strategy used by Reber et al. (1991), Mayberry et al. (1995), and McGeorge et al. (1997), and there is no reason to believe that yet another comparison between implicit and explicit tasks will provide any further insights into the nature of individual differences in implicit learning. An alternative approach to measuring individual differences in implicit learning would be first to assume that both implicit and explicit learning processes can function in a given task, then to attempt to separate the relative contributions of implicit and explicit processes in performance, and finally to account for variability in implicit learning with other measures. On the basis of this reasoning, we developed a new method to measure individual differences in implicit learning.

The task we used was inspired by an early study by Miller (1958), who had subjects learn lists of letter strings in a multitrial free recall experiment. One list contained randomly generated strings, and another list contained strings formed by an artificial grammar. Miller found that subjects could learn the list of grammatical strings considerably faster than the list of random strings: After 10 trials, subjects were able to recall almost nine grammatical letter strings but only three random strings. Miller concluded that grammatical letter strings were easier to learn because they were more redundant and carried less information than random strings. Unfortunately, this is as far as Miller took his artificial grammar learning research program (for a review, see Miller, 1967).

Miller’s (1958) paper is often cited, but his procedure is seldom, if ever, used (but see Redington & Chater, 2002). His original experiment demonstrated nicely that subjects were better at learning and remembering grammatical letter strings than random ones. However, one question that was not addressed in his initial study, and that has that not been explored since that time, is whether an individual who has learned an artificial grammar will improve later memory for new strings generated by that same grammar. Simply put: Can implicit knowledge about an artificial grammar improve immediate memory span for novel sequences generated by that grammar?

The task used in the present experiment was relatively simple and straightforward. Subjects were presented with a sequence of colors and were asked to immediately reproduce the sequence, using a custom-designed response box. Subjects performed the same sequence reproduction task during an “acquisition phase” and a “test phase.” The sequences presented to subjects during the acquisition phase were generated by one of two artificial grammars used in this study (Grammars A and B; see Figure 1), and by reproducing these sequences, subjects were exposed to one of the two grammars. During the test phase, half of the sequences presented were new sequences that came from the same grammar on which the subject had been trained; the other half of the sequences presented during the test phase were new sequences generated by the other grammar that the subject had not been exposed to earlier. The ability to reproduce sequences that came from the *trained* grammar should be facilitated by implicit learning. In contrast, performance on the *not-trained* sequences should not be facilitated by any previous implicit learning, and partialing the not-trained span score from the trained span score should provide a measure of the degree to which implicit learning facilitated immediate memory span performance for each subject.

After completing the acquisition and test phases, subjects were also given a recognition memory test that consisted of several types of sequences. *Old* sequences, selected from the first part of the experiment, came from either the trained grammar or the not-trained grammar. Three types of *new* sequences were also used: sequences that came from the trained grammar, sequences that came from the not-trained grammar, and sequences that were randomly generated and did not conform to the rules of either grammar. At the time of the recognition test, the subject’s ability to recollect old sequences should be quite poor, and they should rely on the familiarity of a given sequence when making their recognition judgments. Sequences that seem more familiar and have increased processing fluency should be judged old more frequently. As a consequence of this “accessibility bias,” which may reflect a form of implicit learning (Jacoby, Debner, & Hay, 2001; Kelley & Jacoby, 2000), subjects should be more likely to call trained sequences old and, more specifically, to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

falsely recognize new sequences from the trained grammar than new sequences from the not-trained grammar.

The sequences of colors were presented to subjects using the Simon memory game, a method of measuring immediate memory span that has been developed in our laboratory (Cleary, Pisoni, & Geers, 2001). In previous research, we have used the memory game to measure immediate memory span in deaf children with cochlear implants, normal-hearing children, and normal-hearing adults. The memory game procedure allows stimuli to be presented to subjects with three different stimulus formats. In the present study, one group of subjects was presented with visual sequences on the memory game (sequences of colored lights); a second group was presented with auditory sequences (sequences of spoken color names); and a third group was presented with multimodal sequences (that is, simultaneous auditory-plus-visual presentation; sequences of colored lights and spoken color names were presented at the same time). After a sequence was presented, subjects were required to reproduce the pattern by pressing the colored response buttons on the memory game box.

We predicted that subjects would have higher immediate memory spans for sequences that came from the grammar they were trained on, compared with sequences that came from the grammar on which they were not trained. We also predicted that subjects would be more likely to call trained sequences old and to falsely recognize novel sequences from the trained grammar, compared with sequences from the not-trained grammar or randomly generated sequences. Such a finding would indicate that subjects had acquired an accessibility bias that interfered with the inability to correctly reject new grammatical sequences. Because this methodology provides an estimate of immediate memory capacity, we should be able to observe a wider range of individual differences in performance than other traditional implicit learning tasks.

METHOD

Subjects

One hundred twenty Indiana University undergraduate students, ages 18 to 24, participated in partial fulfillment of course requirements for introductory psychology. All participants were native speakers of English with no history of speech or hearing disorders and had normal or corrected-to-normal vision at the time of testing.

Materials

Digit span—Tokens of the 10 spoken digits (“0” to “9”) obtained from the Texas Instruments 46-Word (TI46) Speaker-Dependent Isolated Word Corpus (Texas Instruments, 1991) were used for the auditory digit span task. Auditory stimuli were presented over Beyer DT100 headphones calibrated to 74 dB SPL. Subjects recorded their responses by writing in prepared answer booklets at the end of each trial. After making their responses, subjects initiated the next trial by pressing the “enter” key on the computer keyboard.

Simon memory game—Auditory tokens of the four color words (“red,” “yellow,” “green,” and “blue”) were recorded by one male speaker of American English. The memory game response box, modeled after the commercial product “Simon” manufacture d by Milton Bradley, consisted of four colored, backlit response buttons. Subjects recorded their responses to auditory, visual, or auditory-plus-visual stimuli by pressing the response buttons on the memory game.

Two artificial grammars (grammars A and B) were used to generate the grammatical sequences used in the experiment. The grammars were adapted from Brooks and Vokey

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(1991) and are shown in Figure 1. This particular grammar was chosen because it could generate a greater number of short grammatical sequences than other frequently used grammars (see Reber, 1993). Because grammars A and B share the same syntax and the same vocabulary, the sequences they generate should be equally complex and equally difficult to learn and reproduce.

The grammatical sequences used in this experiment were selected pseudorandomly from the set of all possible grammatical sequences, lengths 4 through 10. No sequence with more than three consecutive repetitions of a given color was used, and sequences were selected for use in the acquisition phases so that each branch of the grammar would be represented equally often. No sequence was presented in both the acquisition and test phases.

An additional set of nongrammatical sequences, lengths 5 through 7, was randomly generated for use in the recognition memory test. All random sequences were checked by hand to ensure that they did not conform to the rules of either grammar. These random sequences were allowed to begin with any of the four colors in the vocabulary. However, random sequences beginning with each color were used in equal numbers in the recognition memory test.

Procedure

The subjects were tested in groups of 3 or fewer in a sound-attenuated testing room. All subjects first completed the digit span task, followed by the acquisition phase, the test phase, and the final recognition memory test using the Simon apparatus.

Digit span—The subjects were presented with a list of digits over headphones. Once the entire list had been presented, the subjects wrote down as many digits from the list as they could remember, in the order in which they were originally presented. The lists of digits began at length 4 and increased to length 10, with two lists presented at each length, for a total of 14 trials.

Acquisition phase—The subjects were presented with a sequence of colors using the Simon memory game and were asked to reproduce the sequence by pressing the response buttons. Color sequences were presented either auditorily (a sequence of spoken color names), visually (a visual–spatial sequence of colored lights), or audiovisually (a visual–spatial sequence of colored lights and the same sequence of spoken color names, presented simultaneously). Stimulus presentation format was a between-subjects factor. The subjects were only exposed to sequences that came from one grammar (A or B), and the grammar used during acquisition was counterbalanced across subjects. Sequences began at length 4 and increased to length 10, with two sequences presented at each length, for a total of 14 trials per run. Acquisition consisted of two runs, so that the subjects reproduced a total of 28 different sequences generated by one of the two grammars.

Test phase—The subjects proceeded seamlessly from the acquisition phase into the test phase, without being informed about the existence of the rules that were used to create the sequences. The task during testing was identical to that during acquisition: The subjects reproduced test sequences by pressing the response buttons on the Simon memory game. For each subject, the same stimulus presentation format was used in both acquisition and testing. Half of the sequences used during testing were novel sequences that came from the grammar on which the subject had been trained. The other half of the sequences used in the test phase came from the grammar on which the subject had not been trained. Trained and not-trained sequences were randomly distributed throughout the test phase. Sequences began at length 4 and increased to length 10, with two sequences presented at each length, for a total of 14 trials per run. The test phase consisted of four runs, so that the subjects each reproduced a total of 28 different sequences from the trained grammar and 28 different sequences from the not-trained grammar.

Recognition memory test—After completing the acquisition and test phases, the subjects were told that they would be given additional sequences on the Simon apparatus, some of which they had seen before during the first part of the experiment and some of which were entirely new. If the sequence had been previously presented, the subjects were to indicate *old* by pressing the green button on the Simon apparatus. If the sequence had not been previously presented, the subjects were to indicate *new* by pressing the red button on the Simon apparatus. To help the subjects remember which button corresponded to which response, a small index card labeling each button with the appropriate response was placed on the Simon response box. Once again, the subjects were not informed about the existence of the rules used to generate the sequences prior to the recognition memory test.

The recognition memory test was composed of 50 sequences. Ten old trained and 10 old not-trained sequences were selected from sequences used during the test phase. Ten new trained and 10 new not-trained sequences were generated from each grammar, respectively, and presented as grammatical distractor items. In addition, 10 new random sequences that did not conform to the rules of either grammar were used as distractors. The sequences used in the recognition memory test were of lengths 5 through 7, so that recognition performance would not be confounded by possible capacity limitations in immediate memory span. Two sequences at length 5, four sequences at length 6, and four sequences at length 7 were used in each condition. The sequences in each condition were randomly distributed throughout the recognition memory test.

After the recognition memory test was completed, the subjects were given a postexperiment questionnaire. They were told that the sequences of colors they had been exposed to during the sequence reproduction task were created with a complex set of rules. Subjects were then asked to rate how much they were aware of the rules during the experiment, using a 5-point scale (1 = *not aware* to 5 = *completely aware*), to list any rules or recurring themes that they may have noticed during the sequence reproduction task, and finally to rate how confident they were that they knew the rules, using a 5-point scale (1 = *not very confident* to 5 = *completely confident*).

RESULTS

Memory Span

All memory span scores were obtained by adding the total number of items correctly reproduced on each perfectly recalled trial (an “absolute span score,” after LaPointe & Engle, 1990). This scoring method was chosen because it provides a way of combining both list-based and item-based performance into a single composite score.

The mean memory span scores for the digit span task, the acquisition phase, and the test phase are shown in Table 1, listed by presentation modality group ($n = 40$ in each group). Two scores were obtained during the test phase, one score for sequences from the trained grammar, and one score for sequences from the not-trained grammar. A one-way analysis of variance (ANOVA) performed on the digit span scores revealed no differences among the three presentation modality groups ($F < 1$). The mean scores, ranging from 48.35 to 49.75, are consistent with the findings from other studies carried out in our laboratory using the same methods (Goh & Pisoni, 2003). An additional one-way ANOVA performed on the acquisition phase scores revealed no differences among the three presentation modality groups ($F < 1$), indicating that performance during the acquisition phase was comparable in each presentation modality group.

Sequence type during training (trained or not trained), presentation modality group (AV [auditory-plus-visual], AO [auditory only], or VO [visual only]) and grammar learned during training (A or B) were submitted to a repeated measures ANOVA with repeated measures on the first factor. The analysis revealed a significant main effect of sequence type [$F(1,114) = 73.00$, $p < .001$]. During the test phase, immediate memory span for trained sequences was significantly higher than memory span for not-trained sequences. This “learning effect” is displayed in Figure 2, illustrating that performance on trained sequences was better than performance on not-trained sequences across almost all list lengths. The learning effect did not interact with the grammar (A or B) learned during training [$F(1,114) = 2.38$, $p > .10$]. This finding indicates that the learning effect was not simply a result of sequences from one grammar being easier to learn and remember than sequences from the other grammar. The interaction between the learning effect and presentation modality was marginal [$F(2,114) = 2.68$, $p = .07$].

The size of the learning effect (i.e., trained span score – not-trained span score) for individual subjects is shown in Figure 3. While most subjects showed a learning effect (88 out of 120, 73%), some did not, and there was a wide range of variability in the size of the learning effect among the subjects across all three modality conditions (ranging from −38 to +57). To investigate possible sources of this variability, we calculated correlations of the trained span score, partialing the not-trained span score with digit span and with performance during the acquisition phase. Positive correlations were observed between span scores during the acquisition phase and the magnitude of the learning effect in the AV and AO conditions (*pr*s = +.63 and +.63, respectively; *p*s < .001). In addition, auditory digit span scores correlated with the learning effect in the AV and AO conditions (*pr*s = +.44 and +.47, respectively, *p*s < .01). However, in the VO condition, while a positive correlation was found between performance during the acquisition phase and the size of the learning effect ($pr = +.45$, $p < .01$), no correlation was found between auditory digit span and the size of the learning effect ($pr = .01$).

The finding that performance during the acquisition phase correlated with the size of the learning effect was not entirely surprising—it follows that individuals who performed better while acquiring the grammar would subsequently show a larger learning effect. However, the finding that an unrelated measure of immediate memory span (auditory digit span) correlated with the learning effects in the AV and AO conditions was more surprising. This finding suggests that individuals who have greater immediate memory processing capacity, as measured by auditory digit span, also displayed larger learning effects under the conditions of the Simon sequence reproduction task that also involved auditory presentation.

Recognition Memory

Table 2 shows the probability of calling sequences old in the recognition memory test for trained, not-trained, and random sequences. Examination of Table 2 reveals a highly consistent pattern of results across all three presentation modality conditions. Looking just at the probability of calling new sequences old, there were high false alarm rates for both trained and not-trained sequences, while the subjects seemed to be able to correctly reject new random sequences with relative ease.

The recognition memory data were submitted to a 2 (trained vs. not trained) × 2 (old vs. new) × 3 (AV vs. AO vs. VO) repeated measures ANOVA with repeated measures on the first and second factors. This analysis revealed a main effect of trained versus not trained [$F(1,117) = 26.22$, $p < .001$] and a main effect of old versus new [$F(1,117) = 70.59$, $p < .$

001]. There was no significant effect of presentation modality condition ($F < 1$), and none of the interactions reached significance (Fs < 1). Overall, this pattern of results indicates that subjects in all three modality groups were more likely to call trained sequences old than not-trained sequences, regardless of whether the trained sequences were actually old or new. These findings indicate that increased fluency in processing new trained sequences created an accessibility bias toward calling trained sequences old.

Postexperiment Questionnaire

Immediately following the recognition memory test, subjects completed a postexperiment questionnaire in which they were asked to rate their awareness of the rules during the experiment, list any rules they noticed, and finally rate their confidence that they knew the rules. The mean awareness ratings were 1.63 ($SEM = 0.12$) in the AV condition, 1.43 ($SEM = 0.11$) in the AO condition, and 1.55 ($SEM = 0.13$) in the VO condition. A one-way ANOVA showed no significant differences in awareness ratings among the three groups ($F < 1$). Collapsed across modality conditions, 72 of the 120 subjects indicated that they were *not aware* of any rules (a rating of 1), 34 indicated *slightly aware* (a rating of 2), 13 indicated *moderately aware* (a rating of 3), and 1 subject indicated *completely aware* (a rating of 5).

None of the subjects were able to identify any rules in the form of statements such as, “Green could follow blue, but red could not follow blue.” Only 12 subjects indicated that some colors occurred more frequently than others (e.g., “yellow and red were repeated often”). The majority of subjects who guessed about the rules were not close at all. For example, the one “completely aware” subject wrote, “For every session, you were given 2 or 3 rounds of a certain number of colors. They went in sequence from 4 to 10 or 11.” Fifteen subjects believed that some sequences followed a spatial pattern (e.g., “the colors went around in a circle”). In addition, 5 subjects reported that the particular sequence presented to them on a given trial was determined by their performance on the previous trial (e.g., “The next round usually started with the color I wasn’t sure of ”).

The subjects’ confidence about their knowledge of the rules was also very low. The mean confidence ratings were 1.35 ($SEM = 0.11$) in the AV condition, 1.35 ($SEM = 0.14$) in the AO condition, and 1.23 ($SEM = 0.08$) in the VO condition. A one-way ANOVA also showed no significant differences in confidence ratings among the three groups ($F < 1$). Collapsed across modality conditions, 94 of the 120 subjects indicated that they were *not very confident* that they knew the rules (a rating of 1), 19 indicated that they were *slightly confident* (a rating of 2), 6 indicated that they were *moderately confident* (a rating of 3), and the same subject who earlier indicated *completely aware* also indicated *completely confident* (a rating of 5). Overall, the results of the postexperiment questionnaire demonstrate that even if some subjects claimed to be somewhat aware of the rules, they were unable to explicitly state anything that resembled a rule. Furthermore, even if some subjects attempted to guess about the rules, their verbal explanations obtained after the experiment were not sufficient to explain the robust learning effects observed during the sequence reproduction task.

Multiple Regression Analyses

Our analysis up to this point has not addressed exactly what the subjects learned or what information they used as a basis for responding during the sequence reproduction task or the recognition memory test. Although it is possible that the subjects were learning and using abstract grammatical knowledge (i.e., rules), it is also possible that the subjects were learning and using specific knowledge about exemplars or fragments of sequences that they actually experienced during the acquisition phase (for a review and further discussion, see Shanks, Johnstone, & Kinder, 2002).

To provide a more detailed analysis of the information that the subjects used as a basis for responding in each task, we calculated measures of “associative chunk strength” for each sequence presented during the test phase and during the recognition memory test. Each sequence presented during the acquisition phase, the test phase, and the recognition memory test was partitioned into two- and three-item chunks (bigrams and trigrams). The chunk strength of a given sequence was the average frequency of each chunk in the sequence (see Johnstone & Shanks, 1999; Knowlton & Squire, 1996). The frequency with which each chunk occurred during the acquisition phase was used to calculate the chunk strengths of test phase sequences. The frequency with which each chunk occurred during both the acquisition and test phases of the reproduction task was used to calculate the chunk strengths of recognition sequences.

Sequence length, sequence type (trained or not trained), and chunk strength were entered as predictors in a simultaneous multiple regression analysis using probability of correctly reproducing sequences during the test phase as the dependent variable. Among the three predictor variables, there was no correlation between sequence length and sequence type ($r = .00$), a very small correlation between sequence length and chunk strength ($r = .07$), and a moderately positive correlation between sequence type and chunk strength ($r = .45$, $p < .01$).

The results of the regression analysis, shown in Table 3, revealed a strong influence of sequence length on reproduction performance ($\beta = -.94$, $sr = -.93$, $t = -27.98$), as depicted in Figure 2. More important, sequence type accounted for a small but significant portion of the variance in reproduction performance ($\beta = .09$, $sr = .08$, $t = 2.32$). However, chunk strength did not account for a significant portion of the variance in reproduction performance ($\beta = .03$, $sr = .03$, $t = 0.88$). In fact, the zero-order correlation between chunk strength and probability of correct reproduction was .00. These results suggest that although the subjects did not use chunk strength as a basis for responding in the sequence reproduction task, to a certain extent they did rely on grammatical information in the sequences.

To investigate what factors predicted performance in the recognition memory test, an additional simultaneous multiple regression analysis was performed using the probability of calling a sequence old in the recognition memory test as the dependent variable. Sequence length, sequence type, and chunk strength were entered as predictors. Among the three predictor variables, there was a small correlation between sequence length and sequence type ($r = .04$), a small correlation between sequence length and chunk strength ($r = .07$), and a small positive correlation between sequence type and chunk strength ($r = .21$, $p < .05$). In the regression analysis, shown in Table 4, the influence of sequence length was only marginal ($\beta = -.13$, $sr = -.13$, $t = -1.89$). However, both sequence type ($\beta = .44$, $sr = .43$, $t = 6.09$) and chunk strength ($\beta = .49$, $sr = .48$, $t = 6.82$) accounted for significant portions of the variance in the recognition memory scores. These results suggest that the subjects used a combination of both chunk strength and grammatical information as a basis for responding during the recognition memory test.

DISCUSSION

The primary findings from this study can be summarized as follows. First, we found a robust implicit learning effect in the sequence reproduction task. Immediate memory span for sequences generated by an artificial grammar to which the subjects had been previously exposed was significantly higher than memory span for sequences generated by a different grammar to which the subjects had not been exposed during the acquisition phase. Comparable learning effects were found across all three presentation conditions: AV, AO, and VO.

Second, a wide range of variability in the size of the learning effect was observed among individual subjects. Most of them showed evidence of having benefited from prior exposure to one of the artificial grammars, but a few failed to show any benefit at all. Correlation analyses revealed that the subjects who performed better during the initial acquisition phase, in which they were exposed to one of the artificial grammars, showed larger learning effects during the test phase than did the subjects who performed more poorly. Furthermore, in the AV and AO conditions, significant correlations were found between performance on the auditory digit span task, an independent measure of immediate memory span capacity, and the size of the learning effects observed in the test phase. This finding suggests that individuals who have greater auditory immediate memory spans are better able to acquire implicit knowledge about an artificial grammar in an auditory sequence reproduction task and subsequently use their knowledge to improve their memory span for new sequences that follow the same set of grammatical constraints.

Third, a final recognition memory test revealed that subjects in all three presentation modality conditions were more likely to call trained sequences old than nottrained sequences, regardless of whether the trained sequences were actually old or new. This pattern of results suggests that increased fluency of processing the trained sequences may have created a bias toward incorrectly calling new trained sequences old. Taken together with the results of the sequence reproduction task, the present findings demonstrate a situation in which recall is better than recognition (cf. Tulving & Thomson, 1973): Implicit learning of an artificial grammar facilitated immediate recall of trained sequences but interfered with recognition of trained sequences.

Fourth, our analyses of the postexperiment questionnaire indicated that subjects were generally not aware of the existence of any rules. Even if some subjects claimed to be aware of the rules, they were unable to state anything that resembled rules governing the sequences and were not confident in their guesses. In general, the verbal reports obtained after the experiment cannot account for the robust learning effects we observed during the sequence reproduction task.

Finally, we performed post hoc multiple regression analyses to investigate the sources of information that subjects used as a basis for responding during the sequence reproduction task and the recognition memory test. In the test phase of the reproduction task, sequence type (trained or not trained) accounted for a unique portion of the variance in reproduction performance. However, chunk strength did not. This finding suggests that knowledge about frequently occurring bigrams and trigrams was not responsible for the observed improvement in immediate memory span during the test phase. Instead, grammatical knowledge functioned as the source of information used to produce the observed learning effect. A different pattern of results was found in the recognition memory test: Both chunk strength and grammatical knowledge accounted for significant unique portions of the variance in recognition performance. Overall, these findings suggest that subjects relied on different sources of information in order to accomplish task-specific demands. Grammatical information was used to produce the implicit learning effects observed in the sequence reproduction task, while both chunk strength and grammatical information contributed to recognition memory performance.

A purely abstractionist account of sequence learning cannot explain these findings (e.g., Reber, 1993). If subjects were only abstracting grammatical information from the sequences they were exposed to during the reproduction task, chunk strength should not make any contribution at all to recognition memory performance. Instead, our findings are more consistent with a “detailed episodic coding” framework, which was originally introduced and developed to explain research on perceptual learning of words and voices (Goldinger, 1998; Pisoni, 1997). The detailed episodic coding framework suggests that listeners encode highly detailed information about speech events and that variability in stimulus patterns provides an important additional source of information for the perceptual process. This framework can also be applied to implicit learning. Instead of abstracting out structural regularities and discarding surface information or, alternatively, preserving only surface features and eliminating structural information, the present findings suggest that subjects encoded highly detailed information about specific instances encountered during the sequence reproduction task. Both chunk strength and information about the grammatical structure distributed across the entire set of coded instances was acquired, preserved in memory, and accessed depending on the nature of the task requirements. The conceptual framework for understanding implicit learning in this task is also consistent with the principles of encoding specificity (Tulving & Thomson, 1973) and transfer-appropriate processing (Kolers & Roediger, 1984; Morris, Bransford, & Franks, 1977): Memory performance benefits to the extent that the type of processing operations engaged in during initial encoding are reinstated at the time of retrieval.

The purpose of the present study was not to invent another implicit learning task. In fact, the argument presented here is that there is no such thing as an implicit learning task. Instead, we have explored one possible method of measuring the implicit influences of past experiences on a largely explicit task involving immediate memory span. This novel method of measuring implicit learning in terms of its effects on immediate memory span has also provided new empirical evidence that implicit learning can vary greatly among individuals. The present findings challenge theories of human cognition that do not account for individual differences in implicit cognitive processes (e.g., Reber, 1993, 1997). When implicit learning is measured not by performance on a given task but, instead, as a process that makes some relative contribution to task performance, the variability among individuals in implicit learning processes emerges as a natural consequence of the task-specific demands.

Acknowledgments

This work was carried out as an undergraduate honors thesis by J.D.K. under the direction of D.B.P. This research was supported by NIDCD Research Grant DC00111 to Indiana University. We thank Dale Sengelaub and Tom Busey, who served on the committee, and Miranda Cleary, Luis Hernandez, Lorin Lachs, and Jeff Reynolds for their assistance. The manuscript benefited from comments by Maureen Schmitter-Edgecombe, David Shanks, and one anonymous reviewer.

REFERENCES

Brooks LR, Vokey JR. Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). Journal of Experimental Psychology: General. 1991; 120:316–323.

Cleary M, Pisoni DB, Geers AE. Some measures of verbal and spatial working memory in eight-and nine-year-old hearing-impaired children with cochlear implants. Ear & Hearing. 2001; 22:395–411. [PubMed: 11605947]

Destrebecqz A, Cleeremans A. Can sequence learning be implicit? New evidence with the process dissociation procedure. Psychonomic Bulletin & Review. 2001; 8:343–350. [PubMed: 11495124]

Goh WD, Pisoni DB. Effects of lexical competition on immediate memory span for spoken words. Quarterly Journal of Experimental Psychology. 2003; 56A:929–954. [PubMed: 12881165]

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review. 1998; 105:251–279. [PubMed: 9577239]

Jacoby LL. A process dissociation framework: Separating automatic from intentional uses of memory. Journal of Memory & Language. 1991; 30:513–541.

Jacoby LL, Debner JA, Hay JF. Proactive interference, accessibility bias, and process dissociations: Valid subjective reports of memory. Journal of Experimental Psychology: Learning, Memory, & Cognition. 2001; 27:686–700.

Johnstone T, Shanks DR. Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and Van der Linden (1997). Journal of Experimental Psychology: Learning, Memory, & Cognition. 1999; 25:524–531.

Kelley, CM.; Jacoby, LL. Recollection and familiarity: Process dissociation. In: Tulving, E.; Craik, FIM., editors. The Oxford handbook of memory. Oxford: Oxford University Press; 2000. p. 215-228.

Knowlton BJ, Squire LR. Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. Journal of Experimental Psychology: Learning, Memory, & Cognition. 1996; 22:169–181.

Kolers PA, Roediger HL III. Procedures of mind. Journal of Verbal Learning & Verbal Behavior. 1984; 23:425–449.

LaPointe LB, Engle RW. Simple and complex word spans as measures of working memory capacity. Journal of Experimental Psychology: Learning, Memory, & Cognition. 1990; 16:1118–1133.

Mayberry M, Taylor M, O'Brien-Malone A. Implicit learning: Sensitive to age but not IQ. Australian Journal of Psychology. 1995; 47:8–17.

McGeorge P, Crawford JR, Kelly SW. The relationships between psychometric intelligence and learning in an explicit and an implicit task. Journal of Experimental Psychology: Learning, Memory, & Cognition. 1997; 23:239–245.

Miller GA. Free recall of redundant strings of letters. Journal of Experimental Psychology. 1958; 56:485–491. [PubMed: 13611173]

Miller, GA. Project grammarama. In: Miller, GA., editor. The psychology of communication: Seven essays. New York: Basic Books; 1967. p. 125-187.

Morris CD, Bransford JD, Franks JJ. Levels of processing versus transfer-appropriate processing. Journal of Verbal Learning & Verbal Behavior. 1977; 16:519–533.

Pisoni, DB. Some thoughts on "normalization" in speech perception. In: Johnson, K.; Mullennix, JW., editors. Talker variability in speech processing. San Diego: Academic Press; 1997. p. 9-32.

Reber, AS. Implicit learning and tacit knowledge: An essay on the cognitive unconscious. Oxford: Oxford University Press; 1993.

Reber, AS. How to differentiate implicit and explicit modes of acquisition. In: Cohen, JD.; Schooler, JW., editors. Scientific approaches to consciousness. Hillsdale, NJ: Erlbaum; 1997. p. 137-159.

Reber, AS.; Allen, R. Individual differences in implicit learning: Implications for the evolution of consciousness. In: Kunzendorf, RG.; Wallace, B., editors. Individual differences in conscious experience. Amsterdam: Benjamins; 2000. p. 227-247.

Reber AS, Walkenfeld FF, Hernstadt R. Implicit and explicit learning: Individual differences and IQ. Journal of Experimental Psychology: Learning, Memory, & Cognition. 1991; 17:888–896.

Redington, M.; Chater, N. Knowledge representation and transfer in artificial grammar learning (AGL). In: French, RM.; Cleeremans, A., editors. Implicit learning and consciousness. Hove, U.K: Psychology Press; 2002. p. 121-143.

Shanks, DR.; Johnstone, T.; Kinder, A. Modularity and artificial grammar learning. In: French, RM.; Cleeremans, A., editors. Implicit learning and consciousness. Hove, U.K: Psychology Press; 2002. p. 93-120.

Stadler, MA.; Frensch, PA. Handbook of implicit learning. Thousand Oaks, CA: Sage; 1998.

Texas Instruments. TI 46-word speaker-dependent isolated word corpus [CD-ROM]. Gaithersburg, MD: National Institute of Standards and Technology; 1991.

Tulving E, Thomson DM. Encoding specificity and retrieval processes in episodic memory. Psychological Review. 1973; 80:352–373.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

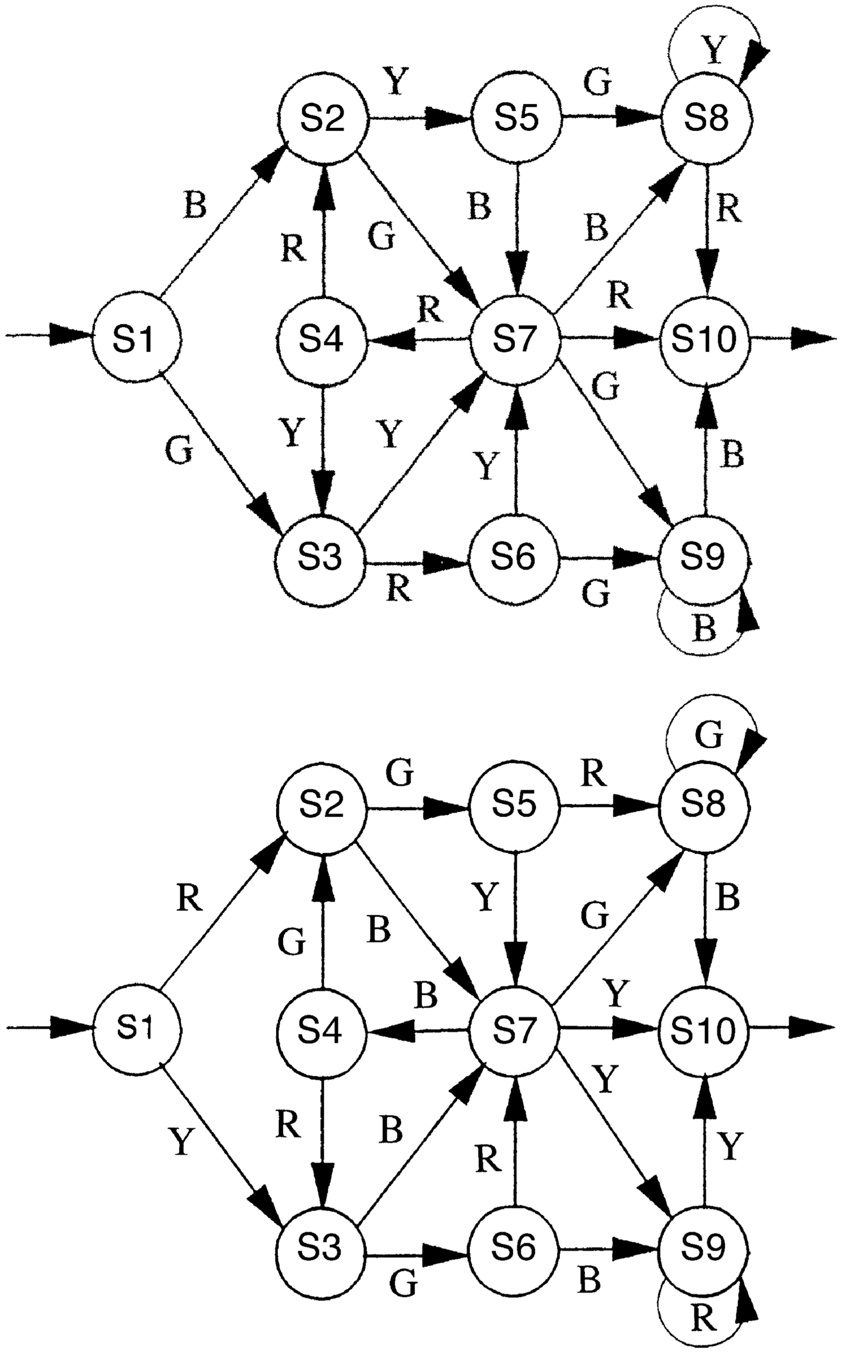

Figure 1.
The two artificial grammars used to generate sequences in this experiment, adapted from Brooks and Vokey (1991). The colors “red,” “yellow,” “green,” and “blue,” are represented as R, Y, G, and B in the diagram. The grammars share the same syntax and vocabulary, but differ in their syntax–vocabulary arrangements.

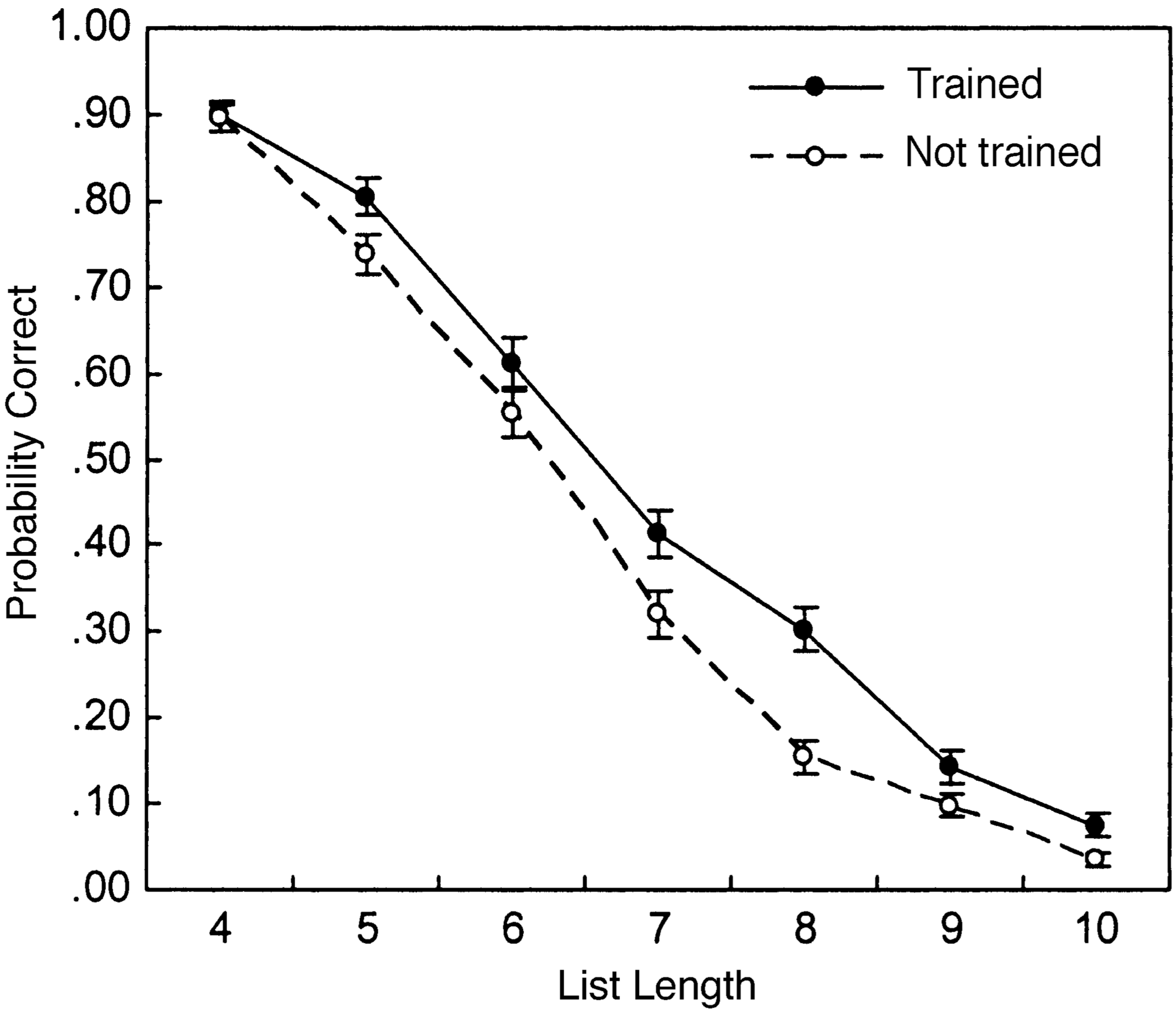

Figure 2.
Depiction of the learning effect (memory span for trained sequences is higher than memory span for not-trained sequences) for all 120 subjects, illustrating that performance on trained sequences is better than performance on not-trained sequences across nearly all list lengths (from list lengths 5 through 10).

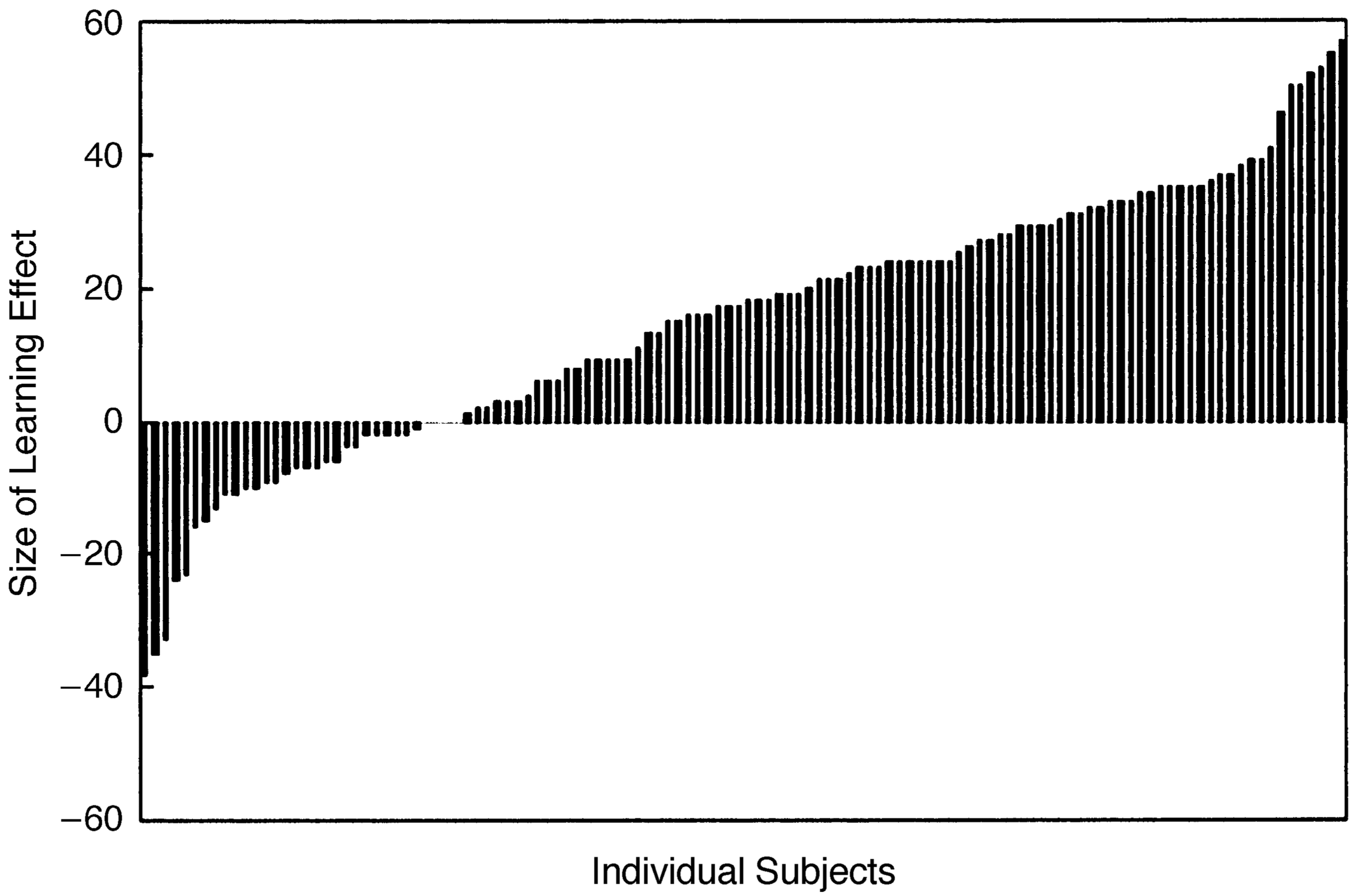

Figure 3.
The size of the learning effect (trained – not trained) for individual subjects rank ordered from smallest to largest effect.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Mean Memory Span Scores

Presentation Modality	**Digit Span**		**Acquisition Phase**		**Test Phase**			
					Trained		**Not Trained**	
	M	***SE***	***M***	***SE***	***M***	***SE***	***M***	***SE***
AV	48.55	3.12	68.98	4.47	83.65	4.20	63.45	3.33
AO	49.75	3.05	64.38	3.63	75.53	4.35	60.88	3.23
VO	48.35	3.07	63.40	4.07	70.90	4.47	60.65	4.27
Total	48.88	1.76	65.58	2.34	76.69	2.53	61.66	2.09

Note—AV, auditory plus visual; AO, auditory only; VO, visual only.

Table 2

Probability of Calling Sequences "Old" in the Recognition Memory Test

Presentation Modality and Sequence Conditions	Sequence Type	
	Old	New
AV		
Trained	.66	.53
Not trained	.60	.47
Random		.25
AO		
Trained	.61	.52
Not trained	.55	.42
Random		.26
VO		
Trained	.62	.55
Not trained	.57	.46
Random		.27

Note—AV, auditory plus visual; AO, auditory only; VO, visual only.

Table 3

Simultaneous Multiple Regression Analysis on Probability of Correct Reproduction With Sequence Length, Sequence Type, and Chunk Strength as Predictors

Variable	*r*	β	*sr*	*t* Value
Sequence length	−.93	−.94	−.93	−27.98**
Sequence type	.10	.09	.08	2.32*
Chunk strength	.00	.03	.03	0.88

Note—R^2 = .88, $F(3,108)$ = 264.75, p < .001; t statistic calculated with df = 108.

* p < .05.

** p < .001.

Table 4

Simultaneous Multiple Regression Analysis on Probability of Calling a Sequence "Old" in Recognition With Sequence Length, Sequence Type, and Chunk Strength as Predictors

Variable	*r*	β	*sr*	*t* Value
Sequence length	−.08	−.13	−.13	−1.89*
Sequence type	.53	.44	.43	6.09**
Chunk strength	.57	.49	.48	6.82**

Note—R^2 = .53, $F(3,96) = 35.42$, $p < .001$; t statistic calculated with $df = 96$.

*$p = .06$.

**$p < .001$.