Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 5, 1002–1011

0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.5.1002

Effects of Model-Based and Memory-Based Processing on Speed and Accuracy of Grammar String Generation

Thomas J. Domangue and Robert C. Mathews
Louisiana State University

Ron Sun
Rensselaer Polytechnic Institute

Lewis G. Roussel
Innovative Emergency Management

Claire E. Guidry
Louisiana State University

Learners are able to use 2 different types of knowledge to perform a skill. One type is a conscious mental model, and the other is based on memories of instances. The authors conducted 3 experiments that manipulated training conditions designed to affect the availability of 1 or both types of knowledge about an artificial grammar. Participants were tested for both speed and accuracy of their ability to generate letter sequences. Results indicate that model-based training leads to slow accurate responding. Memory-based training leads to fast, less accurate responding and highest achievement when perfect accuracy was not required. Evidence supports participants' preference for using the memory-based mode when exposed to both types of training. Finally, the accuracy contributed by model-based training declined over a retention interval.

Mathews, Buss, Stanley, Blanchard-Fields, Cho, and Druhan (1989) proposed that learners draw on two different knowledge sources to guide behavior in complex cognitive tasks. One source is based on their explicit conceptual representation or mental model of the task (e.g., Johnson-Laird, 1982), which is referred to as model-based processing. The second independent source of information is derived from memory-based processing, which automatically abstracts patterns of family resemblance through individual experiences with the task (Brooks, 1978, 1987; Estes, 1986a, 1986b; Hintzman & Ludlam, 1980; Medin & Schaffer, 1978). Because memory-based processing is a byproduct of experience (e.g., encoding exemplars), it does not require intention to learn and often occurs without awareness that anything has been learned (e.g., Reber, 1967). Model-based processing, on the other hand, is effortful (Norman, 1993) and results in a consciously available knowledge that can be readily verbalized. The notion of memory-based processing is consistent with most theories of implicit learning (Knowlton & Squire, 1996; Manza & Reber, 1997; Mathews, 1991; Sun, 2002; Vokey & Brooks, 1992; Whittlesea & Dorken, 1993).

The standard paradigm for studying implicit learning focuses on exposure to sets of instances, usually under instructions to memorize the sets of exemplars (e.g., Mathews et al., 1989; Reber, 1967). In other words, these studies focus on knowledge acquired from memory-based processing alone. Several findings of this body of research have suggested limited usefulness of knowledge acquired from memory-based processing to support complex skills. Some studies (e.g., Dienes & Berry, 1997) have provided evidence that memory-based knowledge is so tied to specific training stimuli that it does not generalize beyond the exact instances experienced during training. Other research has suggested that knowledge acquired through memory-based processing is fragmentary and incomplete (e.g., Dulany, Carlson, & Dewey, 1984; Perruchet & Pacteau, 1990). In addition, research has suggested that people have little confidence in their memory-based knowledge and often feel that they are just guessing when applying this knowledge base (Chan, 1992; Dienes & Berry, 1997).

However, Mathews (1997) argued that these apparent limiting characteristics of memory-based knowledge might be an artifact of the paradigms used to study it. Natural situations that depend heavily on knowledge acquired through memory-based processing (natural language processing or pattern recognition) require extensive practice. Such tasks demand high levels of speed, accuracy, and flexibility. Typical implicit learning experiments involve practice for less than 1 hour. This amount of practice may be inadequate to develop levels of memory-based knowledge that enable accurate and flexible application. In addition, most real-world situations involve a blend of memory- and model-based knowl-

Thomas J. Domangue, Robert C. Mathews, and Claire E. Guidry, Department of Psychology, Louisiana State University; Ron Sun, Cognitive Science Department, Rensselaer Polytechnic Institute; Lewis G. Roussel, Innovative Emergency Management, Baton Rouge, Louisiana.

Claire E. Guidry is now at the Department of Psychology, University of Akron.

A portion of this project served as partial fulfillment of Thomas J. Domangue's master's thesis. Portions of the data were presented at the annual meeting of the Psychonomic Society, New Orleans, Louisiana, November 2000, and at the annual meeting of the Psychonomic Society, Kansas City, Missouri, November 2002. This research was supported by Army Research Institute Grant DASW01–00-K-0012 to Ron Sun and Robert C. Mathews.

We thank Elliot Hirshman, Arthur Reber, and Pierre Perruchet for their helpful comments on a draft of this article. Thomas J. Domangue extends his appreciation to his thesis committee: Robert C. Mathews, Jason Hicks, and Katie Cherry.

Correspondence concerning this article should be addressed to Robert C. Mathews, Department of Psychology, Louisiana State University, 236 Audubon Hall, Baton Rouge, LA 70803. E-mail: psmath@lsu.edu

edge. Thus, it is important to study ways in which these two types of processing interact to influence performance on complex tasks (Sun, Merrill, & Peterson, 2001).

The paradigm used in this series of studies involves having participants generate grammatical strings following different amounts of memory- and model-based processing. To enable our participants to perform this task with a limited amount of training, two letter cues were provided, and 70% accuracy in the generation of strings was considered acceptable. We also limited the time allowed on this cued-generate test; thus, both speed and accuracy were necessary for high levels of performance.

Most theorists accept that some sort of memory of experienced instances (either a neural network, a database of instances, or sets of instance fragments) is the underlying basis for memory-based knowledge (Knowlton & Squire, 1996; Manza & Reber, 1997; Mathews, 1991; Vokey & Brooks, 1992; Whittlesea & Dorken, 1993). However, there are still many questions about what type of training might be optimal for developing such a memory bank of experienced instances.

Some researchers have emphasized the storage of intact exemplars with performance based on the nearest neighbors in the memory bank (Brooks, 1978; Vokey & Brooks, 1992). Hence, a larger database of exemplars should be beneficial when comparing similarities between novel and stored exemplars (Whittlesea & Wright, 1997). Other researchers have proposed that this database contains partial memories of exemplars (Mathews, 1991), memories of chunks of exemplars (Servan-Scheiber & Anderson, 1990), or acquired knowledge of bigrams and trigrams and their frequencies (Perruchet & Pacteau, 1990). This partial memory view might depend more on the representativeness of experienced instances rather than having a large set of instances in memory.

Very little research has examined the effects of mixing memory- and model-based processing. Reber, Kassin, Lewis, and Cantor (1980), in an experiment using a finite-state grammar, found that briefly exposing participants to the actual diagram of the grammar (model-based processing) prior to training with instances (memory-based processing) resulted in better performance on a string discrimination test. In contrast, Mathews et al. (1989) found no advantage of mixed training with a finite-state grammar but did find a beneficial effect of mixed training with a biconditional grammar.

The present series of experiments examines mixing training across sessions as well as an integrated type of training designed to provide simultaneous experience with exemplars (memory-based processing) and knowledge of the structure of the grammar (model-based processing). This new training task is called exemplar diagramming (ExD).

Experiment 1

Two training tasks were contrasted in Experiment 1. One training task, the memory-based or exemplar processing (ExP) task required participants to hold instances in memory long enough to copy them on a response sheet (see Panel A of Figure 1). The other task was integrated training using the ExD task. This task required participants to trace training exemplars through a diagram of the grammar (see Panel B of Figure 1). Thus, participants processed exemplars within the context of the grammar. This experiment also explored the effect of training set size. The small training set

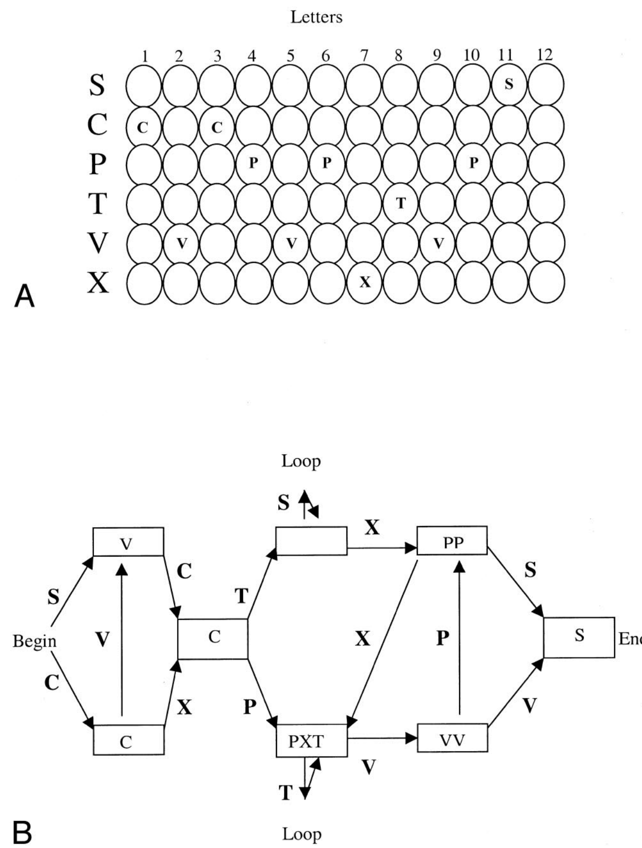

Figure 1. Panel A: The response sheet used to perform the exemplar processing training task. Each letter of a training exemplar is inserted into the circle that intersects the letter to the left and its correct serial position. The valid letter string CVCPVPXTVPS is inserted to illustrate the proper method used. Panel B: The response sheet used to perform the exemplar diagramming training task. The exemplar CVCPVPXTVPS is traced through the map to illustrate proper insertion. The diagram, without the exemplar inserted, was also used for the grammar reproduction training task.

consisted of 22 different exemplars repeated randomly four times (88 instances), whereas the large training set consisted of 88 different exemplars.

Performance was tested using the cued-generate test (Mathews & Cochran, 1997). This test requires generation of a large variety of exemplars based on minimal retrieval cues (two randomly selected letters). We expected that model-based knowledge of the grammar obtained during the ExD task would enhance performance. An explicit representation of the grammar could provide retrieval cues to help access relevant stored exemplars in the memory bank. It could also enhance efficiency and accuracy of string generation by providing a means for correcting errors or omissions in memory traces. In addition, some researchers have suggested that knowledge acquired through only memory-based processing is inflexible (Stadler, Warren, & Lesch, 2000; Dienes & Altmann, 1997). Thus, the memory database created by performing the ExP task might function poorly in enabling generation of diverse sets of exemplars. Therefore, we expected ExD trained participants to outperform the ExP trained participants on the cued-generate test in terms of efficiency (proportion of acceptable strings generated per attempt) and accuracy (number of perfect strings generated). However, using model-based knowledge is known to be a comparatively slow process (Reber et al., 1980; Norman, 1993). Thus, we expected that the ExP group might respond faster than the ExD group and that overall achievement on the test (number of strings generated during a test session) might depend on the optimal balance of speed and accuracy given the task constraints (i.e., 20 min time limit and 70% correct letter match criterion).

Method

Participants

Ninety-two undergraduate psychology students taking a variety of psychology courses at Louisiana State University participated in the experiment. All participants were volunteers and received extra credit for their participation.

Materials

The finite-state grammar used by Mathews et al. (1989) was used in this experiment (see Panel B of Figure 1). This grammar generates 177 exemplars ranging in length from 5 to 11 letters. Two representative subsets of exemplars from this grammar were used as training stimuli. The large subset consisted of 88 exemplars, and the small subset consisted of 22 exemplars randomly repeated four times. Thus, the number of instances was equivalent in both training sets. Each exemplar from both training sets was typed onto the center of a rolodex card and bound to a rolodex base.

In addition, two response sheets were used for the different training tasks: a response sheet used by the memory-based or ExP groups (see Panel A of Figure 1) and a transition diagram of the Mathews et al.'s (1989) artificial grammar used by the integrated training or ExD groups (see Panel B of Figure 1).

Design

The design was a 2 × 2 × 3 (Training Task × Study Set Length × Session) factorial. The two training tasks (ExP vs. ExD) and the length of the exemplar sets (large vs. small) served as between-subjects factors. The three 1-hr weekly sessions served as the within-subjects factor. Twenty-three participants were randomly assigned to each of the four conditions.

Procedure

Participants were tested in groups of up to four. There were three 1-hr sessions scheduled 1 week apart. Each session began with a 20-min training phase requiring participants to perform either the ExP or ExD training task. Each training phase was followed by a 20-min cued-generate test.

Mathews, Roussel, Cochran, Cook, and Dunaway's (2000) cover story was used to provide meaning to the task and motivate the participants. Each participant received a rolodex with one of the two training sets and a packet of response sheets. They were then given a demonstration on how to perform their respective training tasks.

Participants in the ExP groups were instructed to copy as many of the 88 instances (exemplars) as possible into the response sheets in 20 min (see Panel A of Figure 1). Participants in the ExD groups were instructed to trace as many of the 88 exemplars through the diagrams on their response sheets as possible in 20 min (see Panel B of Figure 1). This task was designed to require participants to process exemplars within the context of the grammar's structure.

Testing Phase

The computer displayed two randomly selected letters (cues) and a series of dashes from a not-yet-generated exemplar. Participants filled in the dashes with letters then pressed the enter key. If the letter string generated by the participant did not match at least 70% of the letters of the closest not-yet-generated exemplar, all nonmatching letters were erased and the participant tried again. This process was continued until at least 70% of the letters typed by the participant matched an exemplar. When the 70% criterion was achieved, the computer retrieved the closest not-yet-generated exemplar and displayed it for the participant to observe. Participants then pressed the space bar to begin the next trial with a new test cue.

Because different exemplars may have pairs of letters in common, it was not necessary for the participant to generate the exact exemplar used by the computer to create the two-letter test cue. Thus, participants had some flexibility about which exemplar could be generated on a particular trial. However, once an exemplar was generated, it was removed from the database and could not be generated again during that session. Participants were instructed to generate as many acceptable exemplars (70% or better letter match) as possible during the test and were encouraged to generate perfect exemplars (100% letter match) if possible. All 177 possible strings (ranging in length from 5 to 11 letters) were available at the beginning of each testing session.

Results

A repeated-measures analysis of variance (ANOVA) was used to analyze the data. The results for all four dependent measures are presented in Figure 2. The results on each measure are discussed in turn.

Achievement

Achievement was measured in terms of the number of acceptable strings (70% correct) generated per minute during the 20-min test phase. There was a significant effect of sessions, $F(2, 176) = 262.82$, $MSE = .19$, $p < .001$. Although the achievement levels of all four groups were quite similar (see Panel A of Figure 2), there was a marginally significant effect of list length, $F(1, 88) = 3.41$, $MSE = 1.58$, $p = .068$, and task, $F(1, 88) = 3.56$, $MSE = 1.58$, $p = .063$. Thus, groups with the large training set achieved slightly more than those with the small training set, and groups with the ExP training task achieved slightly more than groups having the ExD training task. The interaction between list length and task was not significant. A Tukey Honestly Significantly Different (HSD) post hoc test of comparisons showed no significant differences between any of the groups on the measure of achievement.

To explore whether more training (old) strings were generated than new (nontraining) strings, the proportion of old hits and new hits from the third session were analyzed. After the third session, participants had maximum exposure to their training set. Therefore, old–new differences should have been maximized in Session 3.

A repeated measures ANOVA showed no significant differences between training set size and no significant differences between the generation of old hits versus new hits. However, there was a significant interaction between training set size and new–old hits, $F(1, 90) = 4.63$, $MSE = .03$, $p = .034$. That is, participants trained with the small set generated significantly more new strings ($M = .304$) than old strings ($M = .263$), whereas participants trained with the large set generated slightly more old strings ($M =$

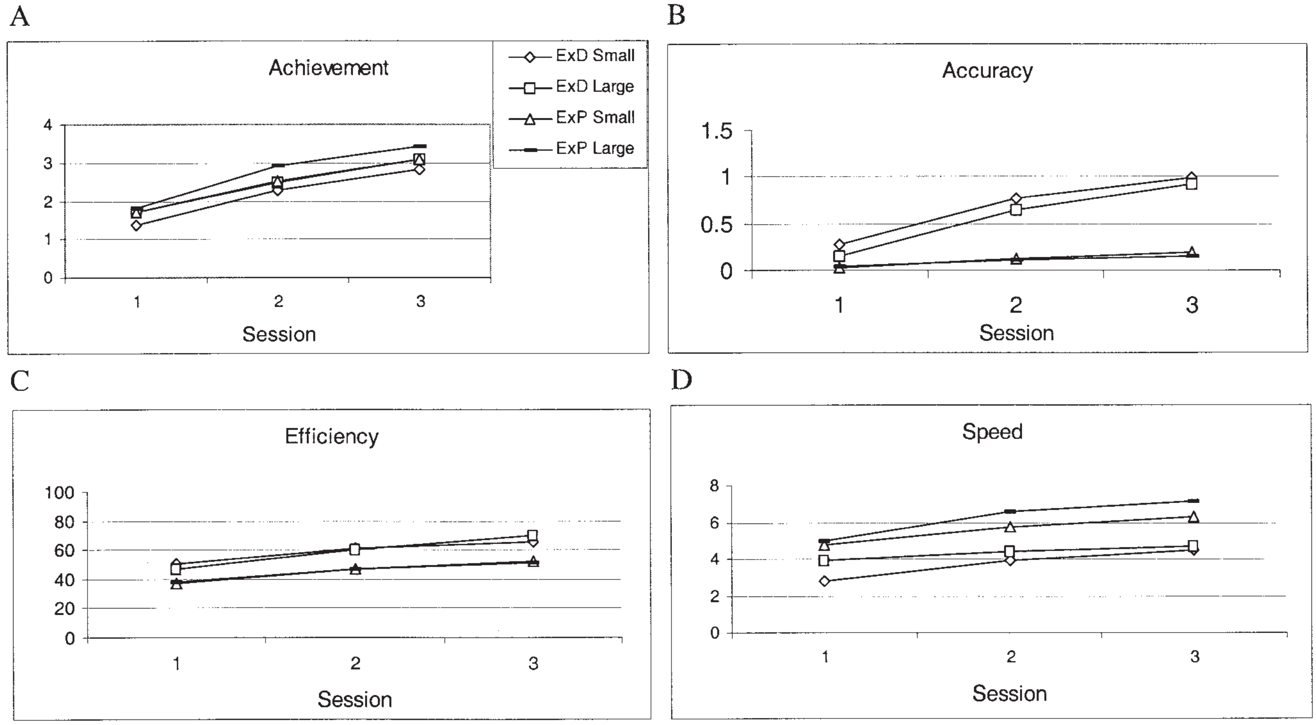

Figure 2. The exemplar diagramming (ExD) and exemplar processing (ExP) training tasks along with a small training set (22 exemplars repeated four times) and a large training set (88 separate exemplars) were used in Experiment 1. Panel A presents scores on the achievement measure. Panel B presents scores on the accuracy measure. Panel C presents scores on the efficiency measure. Panel D presents scores on the speed measure.

.304) than new strings (M = .295). Furthermore, there were no significant differences between ExD training and ExP training on old hits or on new hits.

Accuracy

Accuracy is a measure of the proportion of attempts that matched 100% of the letters in a not-yet-generated exemplar (i.e., the proportion of perfect, 100%, letter strings generated per minute). There were significant effects of sessions, $F(2, 176) = 27.82$, $MSE = .17$, $p < .001$, and task, $F(1, 88) = 17.30$, $MSE = 1.07$, $p < .001$. There was also a significant interaction between sessions and task, $F(2, 176) = 13.23$, $MSE = .17$, $p < .001$. Accuracy of the ExD groups increased more across sessions than did accuracy of the ExP groups (see Panel B of Figure 2).

A Tukey HSD post hoc test of comparisons showed that the small set ExD group ($M = 19.79$) or the large set ExD group ($M = 18.34$) did not differ significantly from each other. However, they both performed significantly better than the small set ExP group ($M = 3.74$) or the large set ExP group ($M = 3.16$), which did not differ.

Efficiency

Efficiency is a measure of the proportion of a participant's attempts that generated acceptable strings. There were significant effects of session, $F(2, 176) = 85.50$, $MSE = 77.67$, $p < .001$, and task, $F(1, 88) = 19.48$, $MSE = 666.54$, $p < .001$. As can be seen in Panel C of Figure 2, the ExD conditions tended to be more efficient than the ExP groups. In addition, all groups became more efficient across the three sessions.

Tukey HSD post hoc tests of comparisons showed that the large set ExD group ($M = 70.16$) performed significantly better than all other groups except the small set ExD group ($M = 65.78$). The small set ExD group only performed significantly better than the large set ExP group ($M = 50.95$). The large set ExP group did not differ from the small set ExP group ($M = 52.59$).

Speed

Speed of responding was measured in terms of the number of attempts (i.e., each time the enter key was pressed) per minute during the test phase. As expected, participants who received model-based training with the grammar (ExD task) responded slower in the cued-generate test than participants who received memory-based (ExP task) training. There were significant effects on sessions, $F(2, 176) = 79.09$, $MSE = .72$, $p < .001$, task, $F(1, 88) = 27.91$, $MSE = 8.92$, $p < .001$, and an interaction between sessions and task, $F(2, 176) = 3.68$, $MSE = .72$, $p = .027$. As can be seen in Figure 2, the ExP groups performed significantly faster than the ExD groups. There was also a three-way interaction between sessions, task, and length, $F(2, 176) = 5.32$, $MSE = .72$, $p = .006$. Whereas the ExP large group increased in speed over sessions more than the ExP small group, the opposite pattern was observed for the ExD groups (see Panel D of Figure 2).

A Tukey HSD post hoc test of comparisons showed the large set ExP group ($M = 7.17$) and the small set ExP group ($M = 6.34$) did not differ from each other but performed significantly better than both the large set ExD group ($M = 4.68$) and the small set ExD group ($M = 4.49$), which did not differ.

Discussion

The results of the first experiment of this series demonstrate that there are both advantages and disadvantages of exposing participants to an explicit representation of the grammar during training. Explicit exposure to the grammar through integrated training in the ExD groups led participants to better accuracy in terms of generating more perfect strings. It also led to greater efficiency in terms of the proportion of strings generated that were acceptable in the cued-generate test (70% correct). However, the ExP groups, who did not have this explicit exposure to the grammar, responded faster, allowing them to generate more acceptable strings during the 20-min test. These results support the view that memory-based knowledge acquired from only exemplar processing is sufficient to support generation of acceptable strings.

There was also a marginal effect of training set size on achievement (number of strings generated). Groups who received the large training set (88 different exemplars) generated slightly more strings than groups who received the small training set (22 exemplars randomly repeated four times). However, this effect was very small, partial $\eta^2 = .036$. In contrast, the groups who were trained using the small training set achieved more hits on new strings than old hits from the training set. Thus, an extensive memory bank of exemplars does not appear to be essential for generating new acceptable strings.

Meulemans and Van Der Linden (1997) also found no significant difference between training set sizes on their generation test. They analyzed their string generation data from Experiments 1B, 2A, and 2B, which had training set sizes of 125, 32, and 125 respectively. Results showed that participants performed equivalent across all three experiments. Hence, Meulemans and Van Der Linden concluded that, "those who had been confronted with a larger number of learning items could not generate more grammatical strings than those who had learned fewer items" (p. 1020).

Experiment 2

In some past experiments, researchers have found that mixing different training tasks across sessions could enhance learning. The present experiment examined the effects of mixing ExP training with ExD training across two weekly sessions. Perhaps groups with mixed training (ExP–ExD or ExD–ExP) would acquire the best qualities of both types of training, faster than ExD only and more accurate than ExP only. Experiment 2 also included a 1-week retention test without a training phase during the third session. This retention test was included because it has often been found that conditions, which lead to the fastest initial learning, do not usually result in the best retention (e.g., Pollock & Lee, 1997; Shewokis, Del Ray, & Simpson, 1998). It was predicted that the group who received ExD training during the first two weekly sessions would perform best in retention because these participants should have retained a mental representation of the grammar in addition to their knowledge of instances acquired from exemplar processing.

Method

Participants

One hundred eight undergraduate students taking a variety of psychology courses at Louisiana State University participated in the experiment. All participants were volunteers and received extra credit for their participation.

Materials

The same materials from Experiment 1 were used in this experiment with the exception of the elimination of the large set of training exemplars.

Design

The design was a one-factor between-subjects design with four levels: ExP during the first two sessions, ExD during the first two sessions, ExP during the first session and ExD during the second session, and ExD during the first session and ExP during the second session. Twenty-seven participants were randomly assigned to each of the four conditions.

Procedure

The procedure was exactly like the first experiment in all aspects except two. First, two groups received a different training task during their second session than they did during the first session (i.e., mixed groups). Second, there were no training tasks during the third session. Instead, everyone performed the cued-generate test for 40 min. The test time was increased in the retention session to obtain a more thorough assessment of participants' ability to generate a wide range of acceptable strings after a 1-week retention interval.

Results

The data from all three sessions are shown in Figure 3. The data from the second session and the retention session are of primary interest because the mixed groups have not experienced both types of training until the end of Session 2.

Acquisition Phase: Session 2 Performance

Achievement. There was no significant effect of training tasks. However, there was a significant effect of the type of string generated, $F(1, 104) = 13.11$, $MSE = .08$, $p < .001$. As in the prior experiment, regardless of group, significantly more new strings were generated ($M = .263$) than old strings ($M = .225$).

Accuracy. There was a significant effect of training tasks, $F(3, 104) = 7.65$, $MSE = .45$, $p < .001$. A Tukey HSD post hoc test of comparisons showed that the ExD–ExD group ($M = .90$) was significantly more accurate than all other groups, which did not differ from each other.

Efficiency. There was a significant effect of training, $F(3, 104) = 5.67$, $MSE = 326.88$, $p < .001$. A Tukey HSD post hoc test of comparisons showed that the ExD–ExD group ($M = 65.88$) was significantly more efficient than all other groups, which did not differ from each other.

Speed. There was a significant effect of training task, $F(3, 104) = 3.69$, $MSE = 4.06$, $p = .014$. A Tukey HSD post hoc test of comparisons showed that the ExP–ExP group ($M = 5.71$) and the ExP–ExD group ($M = 5.57$) performed significantly faster

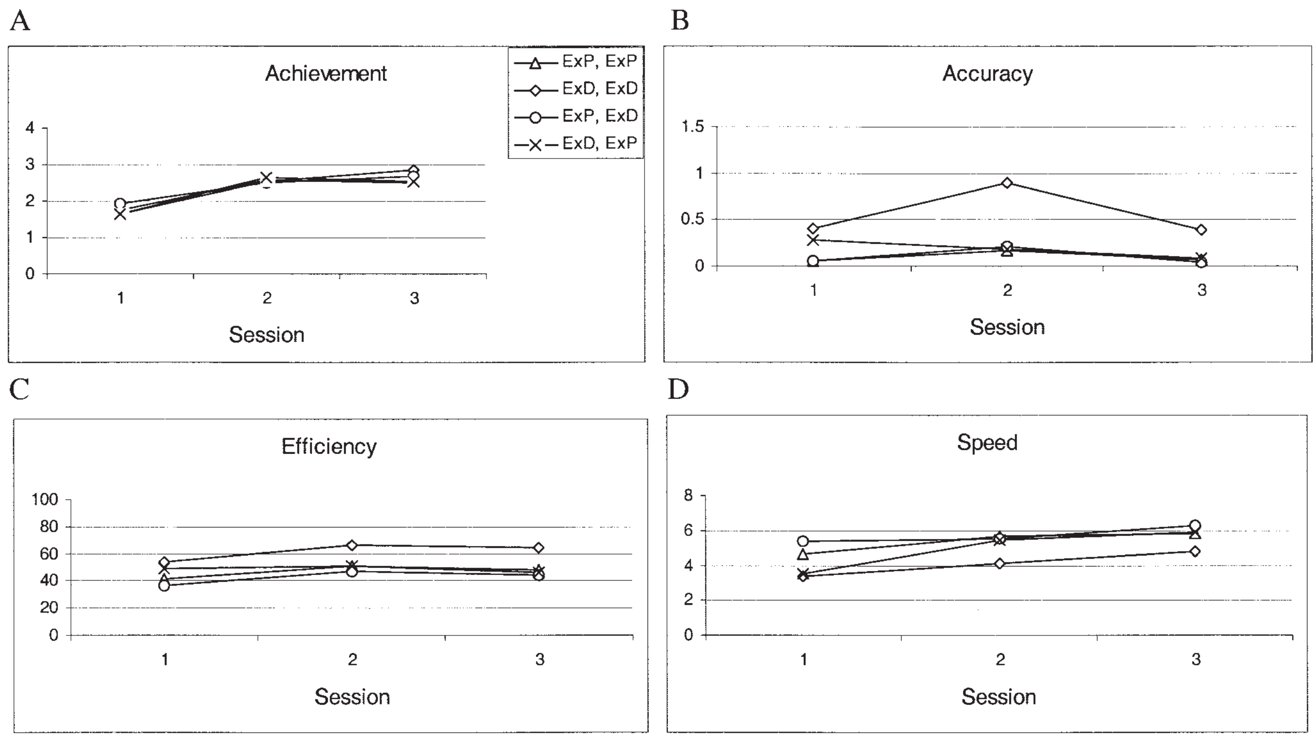

Figure 3. In Experiment 2, the exemplar diagramming (ExD) and exemplar processing (ExP) training tasks were manipulated across the first two weekly sessions. Only the small training set was used. Panel A presents scores on the achievement measure. Panel B presents scores on the accuracy measure. Panel C presents scores on the efficiency measure. Panel D presents scores on the speed measure. The third session contained no training phase and extended the cued-generate test from 20 min to 40 min (testing over a retention interval).

than the ExD–ExD group ($M = 4.11$). The ExD–ExP group did not differ significantly from any other group.

Retention Phase: Session 3 Performance

Achievement. There was no significant effect of the training tasks. All groups maintained their level of achievement on the extended (40 min) retention test. Note that all groups generated approximately the same number of strings per minute in the longer retention session as compared to the 20-min acquisition session (see Panel A of Figure 3). Thus, the rate of generating acceptable strings did not diminish in the extended retention test. As during the acquisition session, significantly more new strings were generated ($M = .544$) than old strings ($M = .459$), $F(1, 104) = 24.24$, $MSE = .39$, $p < .001$.

Accuracy. There was a significant effect of training task, $F(3, 104) = 9.73$, $MSE = .07$, $p < .001$. A Tukey HSD post hoc test of comparisons showed that the ExD–ExD group ($M = .39$) was significantly more accurate after a 1-week retention period than all other groups, which did not differ from each other. However, it should be noted that the ExD–ExD group showed the largest drop in accuracy from Session 2 to Session 3 (see Panel B of Figure 3). This result was surprising because we expected that having both memory- and model-based knowledge of the grammar would enhance retention.

Efficiency. There was a significant effect of training task, $F(3, 104) = 9.73$, $MSE = 234.56$, $p < .001$. A Tukey HSD post hoc test of comparisons showed that the ExD–ExD group ($M = 64.01$) was significantly more efficient after a 1-week retention period than all other groups, which did not differ from each other.

Speed. There was an effect approaching significance of training task, $F(3, 104) = 2.31$, $MSE = 4.48$, $p = .08$. The ExD–ExD group performed slower than all other groups.

Discussion

As in Experiment 1, all types of training led to similar levels of achievement on the cued-generate test during both acquisition and retention. It is interesting to note that the mixed groups performed more like the memory-based processing only (ExP) groups, responding quicker but with less accuracy and efficiency as compared with the ExD groups. This pattern of results suggests that memory-based processing, either before or after model-based processing, led our participants to prefer using the memory-based (fast but less accurate) mode of responding to this task. Perhaps this is because using model-based processing is effortful and slow. In addition, perhaps participants are naturally drawn to the memory-based mode of responding in this task because perfect accuracy was not required (computer motherese was available). Moreover, these patterns were maintained during the 1-week retention interval.

The ExD training task is a mixed form of training. Participants who perform this training task process exemplars (memory-based processing) in the context of a diagram of the grammar (model-based processing). The final experiment of this series adds another training task that is closer to being model-based processing only. This new type of training task, called grammar reproduction or GR, requires participants to commit to memory the diagram of the grammar without processing exemplars during training. In Experiment 3, we also examined mixes of this new more model-based (GR) training task with the memory-based (ExP) training task.

Experiment 3

In Experiment 3, the ExP training task continued to serve as the memory-based processing task, whereas a new training task (GR) was created to provide model-based training only without opportunities to process many exemplars. Very few experiments have provided participants with the grammar diagram during training. In the few studies that have provided such explicit knowledge of the grammar, it was provided for a very minimal amount of time (e.g., Reber et al., 1980). In this experiment, GR trained participants committed the entire diagram to memory before attempting to generate strings.

It was predicted that participants who have memory-based only processing (ExP) would generate strings the fastest. It was expected that the model-based only trained (GR) group would be the most accurate but the slowest. The group that received integrated training (ExD) was expected to fall in between the two single mode of processing groups, using some fast memory-based processes combined with slower model-based processing. We also examined mixed GR and ExP training across sessions to see which type of training produced optimal results for combining memory- and model-based processes. A control group was also added to explore performance in the absence of any type of training task. Although this group had no training, they were expected to perform above chance on the cued-generate test because they could rapidly type each of the six possible letters in succession until a 70% match was obtained. Miller (1969) termed this a cyclic strategy. Thus, the control group might do well in achievement, but their efficiency and accuracy measures were expected to be very low.

Method

Participants

One hundred twenty undergraduate students taking a variety of psychology courses at Louisiana State University participated in the experiment. All participants were volunteers and received extra credit for their participation.

Materials

The same materials used in Experiment 2 were used in Experiment 3.

Design

The design was a one-factor between-subjects design with six levels: ExP during both weeks, ExD during both weeks, GR during both weeks, ExP followed by GR, GR followed by ExP, and a no training control (C) during both weeks. Twenty participants were randomly assigned to each of the six conditions.

Procedure

There were two 1-hr sessions conducted 1 week apart with a 20-min training phase and a 20-min testing phase. Participants followed the same instructions from the prior experiments for performing the ExP and ExD tasks. The GR training task required participants to observe a copy of the artificial grammar for 2.5 min then turn the diagram over. For another 2.5 min, participants reproduced the artificial grammar diagram from memory by drawing it on a blank sheet of paper. This was repeated four times for a total of 20 min training time, consistent with the other training tasks.

The goal of the GR task was to teach an explicit representation of the grammar without showing many valid letter strings that could stimulate memory-based processing. However, it was essential that participants understood how to use the diagram to generate strings. Therefore, prior to the first session, three test cues of increasing complexity were used to demonstrate how to generate strings using the diagram.

The C condition did not receive any training. They were given the six letters of the grammar randomly typed across the middle of a page. These participants' only instructions were to try and generate letter strings by filling in the blanks with combinations of the six letters and to press *enter*. Correct letters would remain on the screen and should have been used in combination with other choices for another attempt until an acceptable string is generated. Participants were also informed about the 70% minimum criterion and the ability of the computer to provide the corrected string.

Testing Phase

The 20-min testing phase was identical to the prior experiments.

Results

Only the results from the second session were analyzed because the mixed groups (ExP–GR and GR–ExP) had not experienced both training tasks until the end of Session 2. However, measures for both sessions are provided in Figure 4.

Achievement

There was a significant effect of training tasks, $F(5, 114) = 6.81$, $MSE = .58$, $p < .001$. A Tukey HSD post hoc test of comparisons showed that the GR–ExP group ($M = 1.34$) and the GR–GR group ($M = 1.34$) performed significantly less well than all other groups except for the C–C group ($M = 1.92$), which did not differ from any group.

As in the previous experiments, there was a significant effect of new hits ($M = .198$) over old hits ($M = .165$), $F(1, 114) = 18.99$, $MSE = .07$, $p < .001$. Unlike the previous experiments, there was a significant effect of training group, $F(5, 114) = 4.07$, $MSE = .06$, $p = .002$. There was also an interaction between type of string generated and training group, $F(5, 114) = 2.77$, $MSE = .01$, $p = .021$.

The ExD–ExD group ($M = .235$) and the ExP–GR group ($M = .235$) generated significantly more new strings than the GR–ExP ($M = .139$) and the GR–GR group ($M = .144$). The C–C group ($M = .208$) did not differ significantly from any group.

Accuracy

There was a significant effect of training tasks, $F(5, 114) = 8.82$, $MSE = .95$, $p < .001$. A Tukey HSD post hoc test of comparisons showed that the GR–GR group ($M = 1.73$) was significantly more accurate than the C–C group ($M = .02$), the

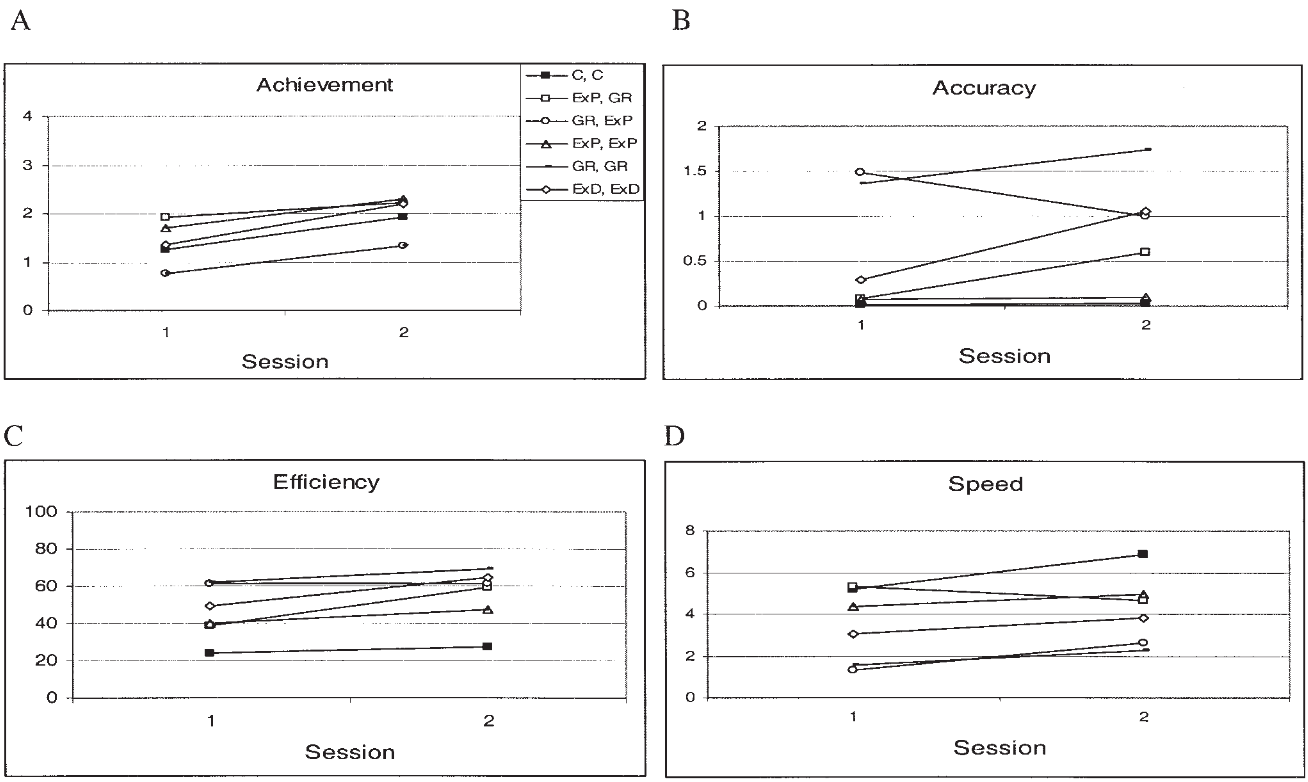

Figure 4. In Experiment 3, three training tasks were manipulated across weekly sessions: exemplar diagramming (ExD), exemplar processing (ExP), and grammar reproduction (GR). There was also a control group (C) that received no training prior to testing. Only the small training set was used. Panel A presents scores on the achievement measure. Panel B presents scores on the accuracy measure. Panel C presents scores on the efficiency measure. Panel D presents scores on the speed measure.

ExP–ExP group (M = .10), and the ExP–GR group (M = .60), which did not differ from each other. The ExP–GR group only differed significantly from the GR–GR group.

Efficiency

There was a significant effect of training tasks, $F(5, 114) = 10.03$, $MSE = 466.50$, $p < .001$. A Tukey HSD post hoc test of comparisons showed that the C–C group (M = 27.54) performed significantly worse than all other groups, whereas the GR–GR group (M = 69.04) was significantly more efficient than the C–C group and the ExP–ExP group (M = 47.41). The ExP–ExP group only differed significantly from the C–C group and the GR–GR group.

Speed

There was a significant effect of training tasks, $F(5, 114) = 14.11$, $MSE = 4.02$, $p < .001$. A Tukey HSD post hoc test of comparisons showed that the C–C group (M = 6.86) was significantly faster than all other groups. The GR–GR group (M = 2.26) and the GR–ExP group (M = 2.61) were significantly slower than all other groups except for the ExD–ExD group (M = 3.83), which only differed significantly from the C–C group.

Discussion

Experiment 3 compared model-based only training (GR) to memory-based training (ExP) and integrated training (ExD). It also examined various mixtures of training type across two sessions. The results followed the pattern of the earlier experiments in that exposing people to a diagram of the grammar (GR or ExD) generally led to slower but more accurate responding on the cued-generate test. Memorizing the grammar without encoding exemplars during training (GR) led to the highest level of accuracy and the slowest responding. Memory-based training led to fast responding with low accuracy. The integrated training (ExD) was in between, having higher accuracy and lower speed than ExP and lower accuracy but higher speed compared with GR.

Whereas, in the earlier experiments, achievement (number of strings generated) was nearly equivalent across groups, in this experiment large differences occurred. The model-based training-only group (GR–GR) had lower achievement, even compared with the control group who had no training. However, the memory-based-only group (ExP–ExP) and the integrated training group (ExD–ExD) were able to generate more strings than the model-based-only group (GR–GR) or the control group (C–C) in Session 2.

It is interesting to note that the groups exposed to mixed training across sessions tended to perform like the single mode of processing groups who had the same type of training in Session 1. Consequently, the GR–ExP group did poorly on the achievement measure (as did GR–GR), and the ExP–GR group successfully generated as many strings as the ExP–ExP group. Thus, it appears that the type of training received initially tends to dominate when training type is changed.[1]

General Discussion

Three experiments contrasted the learning of letter sequences (strings) following various combinations of memory- and model-based processing. Memory-based processing consisted of copying exemplars into a response sheet that required attention to the serial order of letters in each string. Model-based-only processing consisted of memorizing a transition diagram of the grammar. The integrated training consisted of copying exemplars into the transition diagram. A cued-generate test was used to test the ability of participants to generate a wide range of strings under conditions in which perfect performance was not required (70% match to a grammatical string was acceptable).

The overall pattern of results can be summarized very simply: Memory-based training led to fast but relatively inaccurate generation of strings, and model-based training led to very slow but relatively accurate string generation.

The notion that knowledge acquired through memory-based processing is very inflexible was not supported. Groups that only received memory-based training (the ExP task) performed very well on the cued-generate test in terms of total number of strings generated (the achievement measure). In fact, in Experiment 3, the memory-based processing group successfully generated nearly twice as many strings as the model-based processing group (GR).

Experiments 2 and 3 also provided interesting findings concerning attempts to mix the two types of training. In some cases, memory- and model-based processing were switched across two sessions (Experiment 3). In other conditions, memory- and model-based processing were integrated into one type of training exposing participants to exemplars and mapping their structure onto a diagram of the grammar (the ExD task). It is surprising that the integrated training did not lead to greater achievement than memory-based training in any of the three experiments. In addition, in Experiment 2, both groups who received integrated training (ExD) in one of the two sessions and memory-based training in the other (ExP–ExD or ExD–ExP), ended up showing the relatively fast but inaccurate performance associated with only memory-based training. This pattern suggests that participants preferred memory-based responding when exposed to both memory-based and integrated training.

The results were a bit different in Experiment 3 in which memory-based (ExP) training was mixed across sessions with model-based-only (GR) training. Regardless of training order, both of these mixed groups showed the increased accuracy and slower speed associated with model-based processing in Session 2. However, the group who received model-based training first (GR–ExP) did not reach the achievement level associated with only memory-based training or memory-based followed by model-based training (ExP–GR). In some sense, this group received the disadvantages of both types of training: They ended up being relatively slow and inaccurate. It is a bit alarming that this pattern of training might best characterize training outside the laboratory. This would be the pattern associated with formal schooling (model-based training) followed by experience with many cases (memory-based training) when one gets a job.

Memory-based training has the virtue of being relatively effortless, fast, and accurate enough to achieve the most under the present task constraints (20 min with 70% accuracy required). One wonders how much accuracy could be improved with additional memory-based training. Perhaps there is a plateau of performance associated with purely memory-based training that can only be overcome with deliberate (explicit model-driven) practice.

Another question raised by this research concerns the tendency to prefer responding using memory-based processes when exposed to integrated training. Perhaps a similar phenomenon would occur outside the laboratory when people are trained in school with model-based processing and then practice on their job. That is, there might be a tendency to move toward the memory-based mode as one gains experience, and this shift might lead to decreased accuracy of judgment. One recent study of radiologists (Beam, Conant, & Sickles, 2003) supports such a decrease in accuracy in performance associated with practice following completion of formal education. This study found a small but significant drop in cancer detection for each year beyond a doctor's residency training.

The practical messages of this study for training are straightforward: If only accuracy matters, then use model-based training. If only speed counts, then use memory-based training. If both speed and accuracy are important, then the mixed training may be best. However, better results were obtained in Experiment 3 with either the integrated training or memory-based training followed by model-based training (ExP–GR).

One caveat should be mentioned: In Experiment 2, we found the largest drop in accuracy in the integrated training group over a 1-week retention interval. Thus, the advantage of having both memory- and model-based processing may decline quickly over a retention interval. It is not clear if rapid forgetting of the grammar diagram (loss of model-based knowledge) or a shift to using memory-based processing to perform the task causes this effect. Future experiments with a final test of recall for the grammar diagram and reaction time data during the cued-generate test to

[1] We simulated our human data from Experiment 3 with Clarion, an integrative model with a dual representational structure (Sun, 2002; Sun et al., 2001). The model consists of two levels: The top level encodes model-based knowledge, and the bottom level encodes memory-based knowledge. The purpose of the simulation was to see if a model using dual representational structures could capture the key features of our data. The key features we were trying to capture in the simulation were that exposure to a diagram of the grammar either through GR or ExD would enhance accuracy and efficiency, but such exposure would reduce speed. In addition, a high level of achievement could be accomplished through memory-based processing (ExP–ExP) alone, without exposure to a diagram of the grammar. The results, which are available from the authors on request, suggest that the model provides a good fit to these data. Thus, it may serve as a good model for interactions between memory- and model-based processing. The authors thank Xi Zhang for implementing the simulation data.

distinguish use of model-based (slow) memory-based processing (fast) will be helpful in clarifying this issue.

References

Beam, C. A., Conant, E. F., & Sickles, E. A. (2003). Association of volume-independent factors with accuracy in screening mammogram interpretation. *Journal of the National Cancer Institute, 95,* 282–290.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–215). Hillsdale, NJ: Erlbaum.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts in conceptual development: Ecological and intellectual factors in categorization* (pp. 141–174). Cambridge, MA: Cambridge University Press.

Chan, C. (1992). *Implicit cognitive processes: Theoretical issues and applications in computer systems design.* Unpublished doctoral dissertation, University of Oxford, England.

Dienes, Z., & Altmann, G. (1997). Transfer of implicit knowledge across domains: How implicit and how abstract? In D. Berry (Ed.), *How implicit is implicit learning?* (pp. 107–123). Oxford, England: Oxford University Press.

Dienes, Z., & Berry, D. (1997). Implicit synthesis. *Psychonomic Bulletin & Review, 4,* 68–72.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General, 113,* 541–555.

Estes, W. K. (1986a). Array models for category learning. *Cognitive Psychology, 18,* 500–549.

Estes, W. K. (1986b). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General, 115,* 155–174.

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition, 8,* 378–382.

Johnson-Laird, P. N. (1982). Ninth Barlett Memorial Lecture: Thinking as a skill. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 34A,* 1–29.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 169–181.

Manza, L., & Reber, A. S. (1997). Representing artificial grammar: Transfer across stimulus forms and modalities. In D. Berry (Ed.), *How implicit is implicit learning?* (pp. 73–106). Oxford, England: Oxford University Press.

Mathews, R. C. (1991). The forgetting algorithm: How fragmentary knowledge of exemplars can abstract knowledge. *Journal of Experimental Psychology: General, 120,* 117–119.

Mathews, R. C. (1997). Is research painting a biased picture of implicit learning? The dangers of methodological purity in scientific debate. *Psychonomic Bulletin & Review, 4,* 38–42.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1083–1100.

Mathews, R. C., & Cochran, B. P. (1997). Project grammarama revisited: Genereativity of implicitly acquired knowledge. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit knowledge* (pp. 223–259). Thousand Oaks, CA: Sage.

Mathews, R. C., Roussel, L. G., Cochran, B. P., Cook, A. E., & Dunaway, D. L. (2000). The role of implicit learning in the acquisition of generative knowledge. *Journal of Cognitive Systems Research, 1,* 161–174.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Meulemans, T., & Linden. M. V. D. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1007–1028.

Miller, G. A. (1969). *The psychology of communication: Seven essays.* New York: Basic Books.

Norman, D. A. (1993). *Things that make us smart: Defining human attributes in the age of the machine.* Reading, MA: Addison-Wesley.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119,* 264–275.

Pollock, B. J., & Lee, T. D. (1997). Dissociated contextual interference effects in children and adults. *Perceptual and Motor Skills, 84,* 851–858.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning, 6,* 855–863.

Reber, A. S., Kassin, S., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rules structure. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 492–502.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 592–608.

Shewokis, P. A., Del Ray, P., & Simpson, K. J. (1998). A test of retroactive inhibition as an explanation of contextual interference. *Research Quarterly for Exercise and Sport, 69,* 70–74.

Stadler, M. A., Warren, J. L., & Lesch, S. L. (2000). Is there cross-format transfer in implicit invariance learning? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53,* 235–245.

Sun R. (2002). *Duality of the mind.* Mahwah, NJ: Erlbaum.

Sun R., Merrill E., & Peterson T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science, 25,* 203–244.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammar. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 328–344.

Whittlesea, B. W., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General, 122,* 227–248.

Whittlesea, B. W., & Wright, R. L. (1997). Acting adaptively without knowing the consequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 181–200.

Received June 8, 2003
Revision received January 8, 2004
Accepted February 10, 2004 ■