Developing rich and quickly accessed knowledge of an artificial grammar

BILL SALLAS, ROBERT C. MATHEWS, AND SEAN M. LANE
Louisiana State University, Baton Rouge, Louisiana

AND

RON SUN
Rensselaer Polytechnic Institute, Troy, New York

In contrast to prior research, our results demonstrate that it is possible to acquire rich, highly accurate, and quickly accessed knowledge of an artificial grammar. Across two experiments, we trained participants by using a string-edit task and highlighting relatively low-level (letters), medium-level (chunks), or high-level (structural; i.e., grammar diagram) information to increase the efficiency of grammar acquisition. In both experiments, participants who had structural information available during training generated more highly accurate strings during a cued generation test than did those in other conditions, with equivalent speed. Experiment 2 revealed that structural information enhanced acquisition only when relevant features were highlighted during the task using animation. We suggest that two critical components for producing enhanced performance from provided model-based knowledge involve (1) using the model to acquire experience-based knowledge, rather than using a representation of the model to generate responses, and (2) receiving that knowledge precisely when it is needed during training.

Whether on the playing field, on the stage, or in more mundane contexts, one of the most striking aspects of expert performance is the speed and accuracy with which skills are deployed. Although fast and accurate performance may be considered a hallmark of expertise (and a goal of training), it is difficult to study the development of such skills in controlled settings, because it generally takes many years of practice to reach such levels (Ericsson, Krampe, & Tesch-Römer, 1993). Miller (1958) initiated the study of artificial grammars with just this goal in mind; however, he quickly abandoned the idea, because learning an artificial grammar was so difficult and frustrating to learners. With some modifications to the task (Reber, 1967), research continued, and there is a rich literature studying the acquisition of knowledge of artificial grammars (e.g., Knowlton & Squire, 1996; Perruchet, Vinter, Pacteau, & Gallego, 2002; Reber, 1993). However, the aims of this research have differed somewhat from Miller's, in that most work has focused on distinguishing between implicit and explicit learning processes and has not explored combining implicit and explicit learning to bring such knowledge to higher levels of expertise (Mathews, 1997). In fact, these studies typically have involved low levels of exposure to exemplars, and performance on grammaticality judgment tests has often been barely above chance levels. Thus, the participants in these studies have not reached higher levels of gram-

matical knowledge, such as might be required to generate valid strings.

Mathews and Cochran (1998) suggested that two modifications to the artificial grammar paradigm are necessary to study the acquisition of more advanced levels of knowledge. First, it is important that participants have more extensive practice with exemplars of the grammar. Second, the knowledge assessment must involve more than simple discrimination. To address the latter point, they resurrected Miller's (1958) idea of having participants generate valid strings of a complex artificial grammar but modified the task in ways designed to surmount previous problems (e.g., frustration on the part of participants). At test, participants attempt to generate valid strings by performing a *fill-in-the-blank* task in which they are presented with a set of blanks with two letters, or cues, already filled in. Participants attempt to generate letter strings by filling in the blanks, and whenever they match at least 70% of the letters of a previously unproduced valid string, the computer corrects any incorrect letters by providing the full string (a computerized form of "parentese"). Importantly, this *cued generation* test allows for assessments of level of accuracy and speed of production.

Domangue, Mathews, Sun, Roussel, and Guidry (2004) recently used this paradigm to attempt to create highly accurate and quickly accessed artificial grammar knowledge by training participants with mixtures of experience-

R. C. Mathews, psmath@lsu.edu

based and model-based learning (similar to the distinction between implicit and explicit learning; for discussions, see Lane et al., in press; Mathews et al., 1989).¹ Briefly, experience-based learning involves the acquisition of knowledge directly from exposure to exemplars, and model-based learning involves knowledge acquired from a discovered or provided explicit representation of the rules by which the exemplars are generated. In Domangue et al., experience-based training consisted of copying strings onto a response sheet and required attention to the serial order of the letters in the string (exemplar-processing [ExP] condition). Model-based training involved memorizing the diagram of the grammar (grammar reproduction [GR] condition). Integrated training consisted of copying exemplars into the diagram of the grammar (exemplar diagram [ExD] condition), and interleaved training consisted of doing one type of training followed by another (e.g., GR followed by ExP). Following each training session, participants completed the cued generation test, and measures of speed and accuracy were collected.

The results of Domangue et al. (2004) can be summarized very simply: Experience-based training led to fast but relatively inaccurate production, and model-based training led to slow production but more accurate knowledge (i.e., participants generated more 100% accurate strings). Integrated training led to intermediate levels of accuracy and speed. Groups who had a mixture of model- and experience-based training (interleaved) ended up showing the relatively fast but inaccurate performance associated with experience-based learning. Thus, Domangue et al.'s results demonstrate that it is difficult to obtain both fast and accurate performance in artificial grammar learning, even with extensive training. In addition, model- and experience-based processing demonstrated a clear trade-off between high accuracy and speed.

The general goal of the present study was to help illuminate factors that contribute to the efficiency of cognitive skill acquisition. More specifically, in the following experiments, we sought to discover training conditions that are capable of fostering the acquisition of fast and accurate performance in the context of the artificial grammar learning paradigm. The design of these training conditions was based on predictions from current models of artificial grammar learning and on an evaluation of the potential limitations of the training used in Domangue et al. (2004). In each training condition, we emphasized different levels of knowledge (letter, chunk, or grammar structure) as participants learned training strings. To foreshadow our findings, the results reveal that level-of-knowledge emphasis during training does impact participants' ability to produce valid strings quickly and accurately at test, although several additional factors appear to be necessary for such levels of achievement. In the remainder of the introduction, we will discuss the conceptual issues that guided our training conditions and the specific hypotheses underlying our experiments.

One particularly important finding of Domangue et al. (2004) was that high levels of accuracy were obtained only by participants who had studied a representation of the grammar, although these participants produced strings

very slowly. This suggests that acquiring the underlying rules of the grammar is an important aspect of highly accurate performance. Of course, such knowledge may be acquired in different ways, and these paths may affect the speed with which this knowledge can be accessed. For example, most models of artificial grammar learning view it as primarily a bottom-up process in which associations between pairs of letters are acquired first (lower level units), followed by identification of mid-level units, or chunks (e.g., Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). As learning continues, participants may learn about higher level units, such as which chunks occur together and the specific locations chunks can occupy in valid strings. Such models suggest it might be possible to acquire higher level knowledge of the grammar by extensive practice and by facilitating the acquisition of lower level knowledge (because high-level knowledge is "built up" from lower level knowledge through experience). However, other models of learning (e.g., Anderson, 1993; Sun, 2002; Sun, Slusarz, & Terry, 2005) posit that top-down processes may also play a role. For these models, higher level (i.e., structural) grammar knowledge might best be acquired by providing an explicit representation (such as a grammar diagram), along with extensive exposure to exemplars (a combination of model- and experience-based learning). Although these models would not necessarily argue against the possibility of developing higher level representations given enough experience, they do suggest that providing both explicit representations of the underlying rules and experience with exemplars might be a more efficient path to high achievement. For example, achieving high levels of second-language competence likely involves learning that comes from being immersed in a language environment *and* formal training in spelling, grammatical rules, and the proper use of language (Ellis & Laporte, 1997). Both of those approaches to knowledge acquisition suggest that the level of knowledge emphasized during training should impact the ability to acquire accurate and quickly accessed grammar knowledge, although they differ as to which level of emphasis can be predicted to have the greatest utility. These different levels of emphasis are reflected in the training conditions used in both of the present experiments.

Prior research suggests that not all ways of combining explicit representations with experience leads to both fast and accurate performance. Domangue et al. (2004) found that providing participants with a representation of the underlying structure of the grammar at training can enhance accuracy but that production is quite slow. One likely reason is that GR participants had to consult a mental representation of the grammar diagram to guide their generation of exemplars. Furthermore, receiving the diagram and extensive practice "diagramming" exemplars did not lead to highly accurate performance (in fact, accuracy was lower than in the GR condition). Thus, simply practicing with the diagram is not enough to achieve (relatively) expert performance. Anderson (1993) argued that skill acquisition involves transforming declarative knowledge into procedural knowledge. Thus, we might have expected the participants in Domangue et al. who learned by studying the diagram

to become fast and accurate over time as their declarative knowledge of the diagram transformed into procedural knowledge. This was not the case, since the participants in the model-based conditions, although accurate, were very slow all through training. This suggests that these participants had simply learned to rely on an explicit representation of the model to generate strings, rather than using the model during training to acquire procedural knowledge of how to generate strings directly (without needing an explicit representation of the model). Thus, their slow speed was a function of having to use their mental representation of the grammar to generate valid strings.

Two related aspects of Domangue et al. (2004) are likely to have impacted their results. First, their integrated training focused on decomposing training strings into letters for placement in the diagram. Because of this, the participants' attention was primarily on how to use the diagram (learning locations for placement), rather than on the implications of the knowledge contained in the diagram for the construction of good exemplars of the grammar. As will be described below, the diagram conditions in the following experiments involved providing the diagram as a training aid for editing training strings. Thus, the emphasis of the training tasks in these experiments is on creating whole, valid exemplars, rather than on decomposing them on the basis of the rules of the grammar. Another potential factor is that all the training conditions in Domangue et al. were self-paced. It has been argued that rapid decision making is best trained by exposing participants to situations that require judgments under time pressure (Klein, 1998). Thus, it is possible that the GR task did not require the participants to learn the information in ways that would allow this knowledge to be retrieved quickly at the time of test. We note that the training tasks used in all the following experiments required relatively rapid decision making.

In the following experiments, we utilized a computerized string-edit training task (modified from a static version originated by Mathews et al., 1989) to overcome the limiting factors of the previous research described above. In this task, participants saw an invalid string that had errors in it (i.e., letters that were in the incorrect position) and had a limited amount of time to indicate such letters, using a computer mouse. Feedback was provided during the task so that the participants would know which responses were correct (hits/correct rejections) and which were incorrect (misses/false alarms). Training conditions differed in terms of the training aids that were provided on the screen to the participants. In the two experiments reported here, the conditions of primary interest differed in terms of whether the training aids were designed to facilitate the encoding of relatively high-level (in both experiments, a diagram of the grammar) or relatively low-level (in Experiment 1, letters) or mid-level (in Experiment 2, chunks) grammar knowledge. Animation was used to ensure a fast-paced task and to highlight relevant information for each of the conditions. Under these conditions, animation may be helpful because it allows the learners to quickly shift attention from training string to the relevant part of the training cue, obtain the information they need, and return to the string. We note also that this task format

made it nearly impossible for the participants in a diagram condition to try to memorize the diagram and run through it mentally while editing the strings, because such a strategy would have been too slow to keep up with the pace of the string-edit task (see Domangue et al., 2004).

In the present study, we sought to examine the impact of level-of-knowledge emphasis on the ability to acquire accurate and quickly accessed knowledge of an artificial grammar. In both experiments, the participants in all the conditions were given extensive training using the same computerized string-edit training task. Conditions varied with respect to the type of training aids that were available to the participants during the task. Following training, performance was assessed primarily using the cued generation test (e.g., Mathews & Cochran, 1998), which required the participants to produce strings consistent with the grammar. As with Domangue et al. (2004), we were most interested in the participants' ability to produce highly accurate exemplars of the grammar (perfects, or 100% correct responses) and their speed (measured in attempts per minute). In Experiment 1, we compared two hypotheses about the necessary conditions for acquiring accurate and quickly accessed grammar knowledge. First, we sought to test the notion that participants benefit most from receiving high-level grammar information (along with training experience). The participants in our *diagram assistance* condition saw an animated diagram of the grammar while attempting to edit the training string. Letters appeared in their correct position within the diagram, thus providing cues to the participants about the letters to be edited. The animated diagram in this condition provided information to the participants that could be used to correct errors on the presented strings, thus leading them to focus on how the strings were constructed (cf. Domangue et al., 2004). Thus, if a key factor in achieving enhanced performance involves utilizing higher level grammar knowledge, performance in the diagram-assistance condition should be more accurate than and just as fast as that in the control conditions.

An alternative hypothesis was that training that facilitated the encoding of lower, letter-level knowledge would increase the participants' ability to produce highly accurate strings (note that their performance was already expected to be fast, on the basis of the results of Domangue et al., 2004). This was predicated on the notion that it should be possible to prime correct responses on the string-edit task so that the participants would become increasingly knowledgeable about the correct sequences of letters in valid strings. We tested this hypothesis with a *letter assistance* condition in which participants saw a group of letters rise from the bottom of the screen while they were attempting to edit strings. Just prior to a decision regarding a given letter in the string, the participants saw the correct letter, which they could use (rapidly) to make their decision. According to this hypothesis, if processing of lower level grammar knowledge is enhanced, participants should be able to build on this knowledge and acquire the higher level knowledge underlying the grammar; thus, performance should be highly accurate and fast. Performance in the diagram assistance and letter assistance conditions

was contrasted with that in a no-priming control condition (*no assistance*) and in a no-training control condition.

EXPERIMENT 1

Method

Participants and Design

Seventy-two participants were recruited for this study. All the participants were undergraduate students at Louisiana State University and were given extra credit for their participation. Although the participants were assigned at random to conditions, some participants failed to attend the full set of six sessions, leading the groups to be of unequal size: diagram assistance, $n = 17$; letter assistance, $n = 16$; no assistance, $n = 13$; and no-training control, $n = 16$. This study used a mixed factorial design with type of training as a between-subjects factor and type of test as a within-subjects factor.

Materials and Procedure

The finite-state artificial grammar used by Domangue et al. (2004) was used in the present study. The grammar generates 177 letter strings using the letters, S, C, V, X, T, and P. The letter strings in the grammar range in length from 5 to 11 letters.

The participants were tested in groups up to 8. The participants completed six 1-h sessions over the course of 2 weeks (Monday, Wednesday, and Friday of each week). Sessions 1, 2, 4, and 5 consisted of a 20-min training phase and a 20-min cued generation test. In Sessions 3 and 6, the participants completed only a 10-min cued generation test (fixed time test), followed by a 60-trial cued generation test (fixed trial test). In both tests, the participants were given the goal of generating a high number of valid strings. It was expected that implementing a time limit (fixed time test) would place an emphasis on speed, causing the participants to respond quickly but less accurately. Conversely, it was expected that the fixed trial test would result in participants' responding slowly but with greater accuracy, making each attempt "count."

Training phase. Before the training phase began, the participants read a short description of the cover story, which explained that they were learning to make corrections in secret code words. Training involved performing a computerized string-edit task. The participants saw a letter string at the bottom of the computer screen and had to identify the incorrect letters in that string by clicking on those letters with a mouse. Feedback was given during the trial by highlighting correctly marked invalid letters and placing an "X" over valid letters that were marked as incorrect. In addition, at the end of each trial, the game provided feedback regarding the number of misses (incorrect letters that were not identified as such) and false alarms (correct letters identified as incorrect) the participants had made. Instructions encouraged the participants to respond quickly but not to sacrifice accuracy for speed. A monetary prize was offered to the participant who made the fewest errors during training, to further emphasize accuracy over speed. Each time a letter string was displayed, it contained between one and four errors randomly generated by the computer at the beginning of each trial.

Each training session was 20 min in length and consisted of approximately 88 trials. A subset of 22 exemplars from the corpus of 177 was randomly selected for each participant by the computer at the beginning of each session. The same subset was shown four times, each time in a new random order.

To help the participants in their task of identifying errors, the participants in the experimental groups were given different types of training aids to complete the string-edit task. The participants in the no-assistance condition attempted to identify the incorrect letters in the string without any training aids. They were shown a letter string at the bottom of the computer screen and were told to select the incorrect letters in that string and to click on them, using the mouse (see Figure 1A). As the trial progressed, the computer presented the correct string at the top of the screen, with the letters appearing one by one from left to right until the entire string

Figure 1. Screen shots of the training conditions in Experiment 1. (A) The no-assistance condition. (B) The letter assistance condition. (C) The diagram assistance condition.

was revealed. Approximately 3 sec after the start of the trial, letters began appearing at the top of the screen at a rate of 500 msec per letter; 500 msec before a letter appeared in its position at the top of the screen, the participants could no longer edit the letter in that position. Thus, the participants were forced to make a choice about the validity of each letter (and mark incorrect letters) in the to-be-edited string *before* its corresponding letter became visible at the top of the screen.

In the letter assistance condition, the participants were given the same string-edit task as in the no-assistance condition but were provided with an aid to prime correct choices (see Figure 1B). Instead of the correct letters' appearing one by one at the end of the trial, as in the no assistance condition, the letters began the trial at the bottom of the screen in an unreadable "bunch," where all the letters were drawn on top of each other. Approximately 2.5 sec into the trial, the first letter became recognizable and could be seen approximately 500 msec before it passed a line drawn across the middle of the screen. Similarly, each additional letter crossed the line at a 500-msec interval. After the letters had passed this line, the participants could no longer mark the corresponding letters in the to-be-edited string as incorrect. Thus, the correct response was primed in the sense that the correct letter could be seen approximately 500 msec before time was up for editing the corresponding letter in the training string. The participants could see the correct "answer" for each letter in the to-be-edited string and could mark any invalid letters.

The participants assigned to the diagram assistance condition received the same string-edit task as in the other conditions but were provided with a diagram of the grammar for assistance (see Figure 1C). Instead of the letters floating to the top of the screen, as in the letter assistance condition, the letters appeared, one by one, in the correct order and position in the diagram. The first letter appeared in the diagram 2.5 sec into the trial, and letters appeared at 500-msec intervals. After a letter had appeared in the diagram, the participants could no longer click on the corresponding letter in the string, forcing the participants to predict which letter would appear next.

Cued generation test. After the 20-min training phase in Sessions 1, 2, 4, and 5, the participants read the instructions for the cued generation test. The participants were instructed that their task was to generate as many secret code words (exemplars) as possible during the 20-min test period. They were also told that they should

work as quickly and as accurately as possible and that the participant who found the most code words would receive a monetary prize. Rewarding the total number of valid strings generated is, in effect, rewarding both speed and accuracy. The participants in the no-training control condition were given a sheet of paper with the letters from the grammar printed horizontally across the page in a random order, since they had no prior experience with the grammar and, thus, did not know which letters could be used. They were told that they were to generate secret code words based only on the letters on the sheet of paper. They were also told that the code words would not be English words (e.g., the letter set used in this grammar did not contain vowels).

At the beginning of each test trial, the computer randomly selected a target exemplar (see Figure 2). A set of dashes was displayed on the screen, with one dash for each letter in the target. Two letters (the cues) were displayed in their correct position above the appropriate dashes. Across trials, the letter cues were randomly assigned to different ordinal positions within the target. Working from left to right, the participants typed letters, one for each dash, and pressed the Enter key. The computer then compared the letters that the participant had entered with all of the not-yet-generated exemplars in the database. If the string entered by the participant did not match at least 70% of the letters in a valid exemplar, the computer erased any incorrect letters, leaving only letters that matched the closest valid exemplar left in the database. The participants continued this process until at least 70% of the letters matched a not-yet-generated exemplar. A number of exemplars within the corpus had letters in common, so it was not necessary that the participant should type the target exemplar chosen by the computer but *only that it should match an exemplar not yet generated during the test*. When the participant reached the 70% criterion, a feedback screen appeared in which the letter string generated by the participant was displayed along with

Figure 2. An example of five successive screens in the cued generation test used in both experiments.

the closest matching exemplar. Once 1 of the 177 exemplars had been produced, it was removed from the database and could not be generated again until the next test session.

No training phase was administered in Sessions 3 and 6. Instead, the participants completed two versions of the cued generation test. In the first version, the fixed time test, the participants were given 10 min to generate as many exemplars as possible. They were instructed that they could make an unlimited number of attempts, defined as each time the participants typed letters over the dashes and pressed the Enter key. In the fixed trial version of the test, the participants were told there would be no time limit but that they would be allowed only 60 attempts, so they should be as accurate as possible. The goal of these tests was to emphasize speed and accuracy, respectively. However, as will be detailed below, the pattern of differences between the groups did not differ for any of the dependent measures (i.e., there were no significant training $\times$ test interactions).

Results and Discussion

We will present the test data from the sixth and final session, where the effect of the manipulation would be expected to be strongest. The data from the third session show a pattern similar to that for the sixth session, although some effects were smaller. There were no group $\times$ session interactions. The group means for all dependent measures are presented in Table 1. For each measure, we conducted a 2×4 mixed factorial ANOVA with type of training (diagram assistance, letter assistance, no assistance, or no-training control) as a between-subjects factor and type of test (fixed trial or fixed time) as a within-subjects factor. Unless otherwise noted, the results were significant at the .05 alpha level.

Acceptable Strings

We first examined how often the participants were able to produce a string that met a moderate criterion for accuracy (70% correct) on their first attempt. We measured this as the number of letter strings generated on the first attempt that matched at least 70% of the letters in the target exemplar over the total number of first attempts (see Table 1). Analyses revealed a main effect of type of test, with participants producing more acceptable strings in the fixed time test ($M = .50$) than in the fixed trial test ($M = .45$) [$F(1,57) = 9.56$, $MS_e = .009$, $\eta_p^2 = .14$]. Most important, there was also a significant main effect of training [$F(3,57) = 4.24$, $MS_e = .16$, $\eta_p^2 = .18$]. Post hoc comparisons revealed that all the training groups showed better per-

formance than did the no-training control group. No other comparisons were significant. Finally, the test $\times$ training group interaction was not significant ($F < 1$). Thus, when accuracy is assessed in terms of acceptable strings, there are clear advantages of every type of training over the no-training control. However, 70% correct is a fairly modest criterion for accuracy, which could be accomplished even if a participant had only partial knowledge of the grammar. As in previous work (e.g., Domangue et al., 2004), we examined the participants' ability to produce highly accurate (100%) exemplars of the grammar.

Perfect Strings

In terms of our hypotheses, we were most interested in whether training affected the ability to produce strings that met a very high criterion for accuracy. Using this higher criterion of accuracy allowed us to differentiate between adequate performance (70%) and truly expert accuracy. To do this, we measured the number of perfect strings (100% correct match) generated on the first attempt over the total number of first attempts (see Table 1). Analyses revealed a main effect of type of test, with participants producing more perfect strings on the first attempt on the fixed time test ($M = .12$) than on the fixed trial test ($M = .10$) [$F(1,57) = 5.98$, $MS_e = .002$, $\eta_p^2 = .10$]. Most important, we obtained a main effect of type of training [$F(3,57) = 5.8$, $MS_e = .06$, $\eta_p^2 = .23$]. Post hoc comparisons revealed that the diagram assistance condition had produced significantly more perfect strings than had all the other conditions. No other comparisons were significant. The interaction of test and type of training was not significant ($F < 1$). Thus, the participants who performed the string-edit task while assisted by the animated diagram appeared to have exceptionally good knowledge of the grammar, since they produced perfect strings on an average of 25% of their first attempts.

Speed

Studying the development of expertise requires examining both accuracy and speed. Speed of performance was measured as the number of attempts per minute. This measure cannot be viewed in isolation, since high levels of speed with low accuracy would be like a novice second-language learner randomly verbalizing phonemes in the

Table 1
Performance on the Cued Generation Test in Experiment 1

Measure	Condition							
	DA		LA		NA		Control	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Fixed Trial Test								
Accuracy—acceptable	.55	.33	.49	.49	.52	.21	.22	.24
Accuracy—perfect	.25	.24	.09	.10	.04	.04	.01	.02
Speed	8.00	2.52	8.24	2.83	7.80	1.67	10.11	2.67
Fixed Time Test								
Accuracy—acceptable	.59	.33	.54	.32	.58	.26	.29	.28
Accuracy—perfect	.27	.32	.10	.11	.07	.06	.02	.03
Speed	8.01	2.99	7.10	2.10	8.26	1.80	9.66	2.12

Note—DA, diagram assistance; LA, letter assistance; NA, no assistance.

hope of generating a valid sentence. Thus, it is important to demonstrate that the improved accuracy found in the diagram assistance condition is not limited to a speed–accuracy trade-off. If there is no trade-off, we would expect that the diagram assistance condition would be at least as fast as the other conditions.

The analysis revealed that the type of test did not have a significant impact on speed ($F = 1.5$). Thus, the expectation that the fixed time test would promote speed and the fixed trial test would encourage accuracy was not confirmed. There was also no significant interaction of test and type of training [$F(3,57) = 3.44$, $MS_e = 1.6$, $\eta_p^2 = .10$]. There was a main effect of type of training [$F(3,57) = 3.2$, $MS_e = 9.99$, $\eta_p^2 = .14$], with subsequent comparisons revealing that the no-training control group made significantly more attempts than did all the other groups. No other comparisons were significant. The finding of faster responding in the no-training control group was not unexpected, because participants without training often adopt the strategy of quickly typing in any of the six possible letters (see, e.g., Domangue et al., 2004) and taking multiple trials to reach the 70% criterion. As was noted above, these control participants were significantly less accurate than those in the training conditions, even using a moderate criterion to assess accuracy. No other comparisons were significant. For the purposes of this experiment, the most important finding was that diagram assistance participants were just as fast as participants in the other training conditions and, as was noted above, were also highly accurate.

The results of Experiment 1 were clear and straightforward. All the training groups showed similar rates of production at test and proportions of acceptable strings on their first attempt. However, in line with our first hypothesis, diagram assistance participants demonstrated more detailed knowledge of the grammar, since they produced significantly more perfect strings on their first attempts than did those in the remaining conditions and were just as fast. Thus, our findings suggest that conditions that enhance the processing of higher level grammar knowledge during training are capable of producing both fast and accurate performance, although how such knowledge is introduced plays a major role in its effectiveness (see Domangue et al., 2004).

We also found that directly priming correct letter responses of participants during training did not significantly enhance grammar knowledge. The letter assistance participants showed the same fast and moderately accurate performance as did the participants who had not received priming assistance during training. Thus, at least within the constraints of this manipulation, our second hypothesis was not supported. Priming responses in a letter-by-letter fashion during training does not appear to enhance the acquisition of higher levels of grammar knowledge needed to support highly accurate performance.

EXPERIMENT 2

In contrast to prior work by Domangue et al. (2004), Experiment 1 demonstrated that it is possible to achieve high levels of accurate grammar knowledge while maintaining high rates of production using an animated diagram to as-

sist participants during training. However, priming letter-level knowledge through our animated letter assistance condition did not facilitate highly accurate generation of strings. One influential viewpoint in the artificial grammar learning literature argues that improved performance occurs primarily through increased efficiency at analyzing and identifying exemplars with respect to chunks (Peruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). Thus, the letter-level emphasis may not have assisted and may, potentially, have discouraged the process of building up grammar knowledge in a bottom-up manner. Instead, participants may be able to better engage such processing when knowledge is emphasized at mid-levels (chunks). In Experiment 2, we contrasted our animated diagram condition (providing high-level information) with chunking-assisted groups who received information about how the strings could be consistently processed in chunks. By guiding this step in bottom-up learning, we attempted to facilitate the acquisition of highly accurate grammar knowledge without providing top-down information (i.e., the higher level rules of the grammar).

In Experiment 2, we also sought to ascertain precisely which aspects of the animated diagram are responsible for helping achieve high levels of performance by utilizing a design that allowed us to test between several salient explanations of this effect. We had several competing hypotheses with respect to the critical factor(s) underlying performance in the diagram assistance condition in Experiment 1. The primary hypothesis was that the high level of performance observed in the diagram assistance task was a function of an interaction between animation and the presence of the diagram. One important aspect of the diagram assistance task used in Experiment 1 is that the animation highlights the appropriate part of the diagram just when participants are editing the corresponding part of the exemplar. This hypothesis suggests that performance should be best in conditions that combine the use of the diagram with animation, because it provides the critical information just when it is needed to edit each part of a string. We examined this factor by comparing an animated (just-in-time) version with a static (all-at-once) version of the diagram.

In a related vein, we also sought to test another hypothesis regarding the effectiveness of diagram assistance training. Given that this task encouraged the participants to predict which letters would appear next in the diagram, one possibility is that this led to a type of generation effect (e.g., Slamecka & Graf, 1978), leading to better memory for parts or whole exemplars. This hypothesis suggests that conditions that require participants to predict the relevant letter in the training string should later produce more highly accurate strings on the cued generation test than will conditions in which correct letter information is available at the time of edit task (which allows the participants to compare the provided information with the training string to make a judgment). To test this hypothesis, we compared groups who received the information just before or just after they had to finish editing a particular letter in a string. Thus, the immediate (before) group did not have to generate the required information, but the predictive (after) groups had to generate the letter before it appeared.

In Experiment 2, we used a $2 \times 2 \times 2$ factorial between-subjects design and manipulated the following aspects of the string-edit task: display content (chunk vs. diagram), display type (animated vs. static), and prediction (predictive vs. immediate). This design allowed us to test the hypotheses noted above and also to rule out simpler and more complex alternative hypotheses. In addition to the cued generation test used in Experiment 1, we assessed knowledge of the grammar by having the participants complete a grammaticality judgment test on the last session. This test is commonly used in artificial grammar research and involves having participants judge whether a given exemplar is or is not valid with respect to the grammar. On this test, we also varied the types of invalid exemplars the participants had to judge. Specifically, one type of error involved exemplars that were invalid because one or more letters within a chunk could not occur in a given position (*within-chunk errors*). The participants would need to have good knowledge of dependencies within a chunk to accurately identify such exemplars as invalid. A second type of error involved exemplars that were invalid because a chunk was in a wrong position (*chunk position errors*). The participants would need to have good higher level knowledge of the grammar to accurately identify such errors. These error types allowed us to more precisely assess the type of grammar knowledge acquired by the participants during training and to relate this knowledge to performance on the cued generation test. Given that the test type manipulation failed to produce the desired result in Experiment 1, only a fixed time test was used in Experiment 2.

Method

Participants and Design

One hundred eighty-seven participants were recruited for this experiment. All the participants were undergraduate students at Louisiana State University and were given extra credit for their participation. Although the participants were assigned at random to conditions, some participants failed to attend the full set of five sessions, leading the groups to be of similar but unequal size (*ns* ranged from 19 to 22). This study used a $2 \times 2 \times 2$ factorial design, with content (diagram or chunks), presentation type (static or animated), and prediction (predictive or immediate) as between-subjects factors. A no-training control condition was also included.

Norming Study

To determine the appropriate chunks for the chunk-training condition, 25 undergraduates were recruited for extra credit in their psychology courses. The participants had no prior experience with artificial grammar learning experiments or with the particular grammar used. The participants were given a packet of the 177 exemplars of the grammar, with approximately 30 exemplars printed on each page. The exemplars were arranged in a random order. The participants were told to read the letters in the exemplars from right to left and to place a line between letters where they naturally paused, just as they would pause after clauses while reading a sentence. The participants completed the first packet and then were given a second packet. The second packet contained all 177 exemplars in a different order from that for the first packet. The participants were given the same instructions. Prior work (Perruchet et al., 2002) had shown that segmentation of exemplars became consistent across participants with increased exposure, so only data from the second packet were used.

Each division the participants made was considered a chunk, and the frequency of each chunk was analyzed. The goal was to develop the smallest set of chunks that could, in various combinations, pro-

duce any exemplar in the corpus. With this in mind, the most frequently generated chunks from the pilot study were identified. From that list, the smallest set (17) of chunks that could generate the entire corpus of exemplars was selected (see the Appendix).

Procedure

The participants were tested in groups up to 8. The participants completed five 1-h sessions over the course of 1 week. Sessions 1–4 consisted of a 20-min study phase and a 20-min cued generation test. In Session 5, the participants completed a 20-min cued generation test, followed by a grammaticality judgment test. The same finite-state artificial grammar as in Experiment 1 was used in this experiment.

Training phase. Each training session was 20 min in length and consisted of approximately 88 trials. The entire corpus of 177 exemplars that can be generated from the grammar were utilized in this study. As in Experiment 1, in any given session, the participants were trained on a set of 22 exemplars. The set from which these were drawn was as follows. First, a set of 22 exemplars that were representative of the whole corpus (see Domangue et al., 2004) was divided into two subsets of 11 exemplars each, Sets A and B. The sets were rotated so that Sessions 1 and 3 used Set A and Sessions 2 and 4 used Set B. Thus, these exemplars were repeated in two sessions, and the participants saw 11 of this type at any one session. We also note that these repeated items served as targets in an episodic recognition test in Session 5. This episodic test was included in order to examine the relation between episodic recall and grammar knowledge. The results are not reported because there was a floor effect: There was no evidence of above-chance episodic knowledge in any of the conditions. A separate set of 22 exemplars was withheld (not included in training or the cued generation test) for later use as lures in the episodic memory test. Second, in each session, a set of 11 exemplars was randomly selected from the remaining 133 exemplars. Thus, in any given session, the participants saw 11 exemplars from the repeated set (A or B) and 11 randomly selected exemplars. The order in which the exemplars were displayed in the string-edit task was determined by the computer, using random selection without replacement. When each exemplar had been displayed once, the computer generated a new order. The participants viewed each exemplar approximately four times during the study phase. Over the course of the four training sessions, the participants viewed each of the representative exemplars eight times.

Training involved a modified version of the computerized string-edit task from Experiment 1. As in Experiment 1, the participants were given training aids for completing the string-edit task. Each training condition was characterized in terms of values on three dimensions: display content (diagram vs. chunk), display type (animated vs. static), and prediction (predictive vs. immediate). Figures 3 and 4 provide example states from the diagram-animated predictive and chunk-animated predictive tasks, respectively.

In the diagram-static immediate condition, the participants saw a diagram of the grammar used to generate the exemplars in the middle of the screen. At the beginning of each trial, all of the letters in the exemplar were shown in their appropriate state in the diagram. The participants could then compare the letters in the diagram with the letters in the string they were editing and mark errors where appropriate. Three seconds after the trial began, a dot appeared under the first letter in the to-be-edited string. After 500 msec, the dot moved to the next letter in the to-be-edited string, and the participant was no longer allowed to mark the first letter as incorrect. After another 500-msec interval, the dot moved to the third letter in the string, and the participant was no longer allowed to edit the second letter, and so on. Thus, the dot was a visual timing device that forced the participants to make quick decisions in the edit task. After the dot had passed a letter, an unmarked error was recorded as a miss, and an “X” mark appeared over that letter, alerting the participants to their error. False alarms were also marked with an “X.” The timing dot and “X” marks for errors were consistent across all the conditions. In addition, the time constraint for responses of 500 msec per letter

Figure 3. Example states from the diagram-animated predictive condition in Experiment 2.

was also consistent across conditions. At the end of each trial, the correct exemplar was displayed at the top of the screen.

In the diagram-static predictive condition, the participants performed the string-edit task while the diagram was displayed on the screen. At the end of each trial, all of the letters from the exemplar appeared simultaneously in their appropriate state within the diagram and remained for 3 sec. Thus, the display did not change (it was "static") during the trial. Because the letters did not appear within the diagram until the end of the trial, the participants had to predict which letters would appear in the diagram in order to correct the to-be-edited string.

The participants in the diagram-animated immediate condition saw the letters appear one by one in the diagram. After a letter had appeared in the diagram, the participants had 500 msec to compare that letter with the corresponding letter in the string they were editing and to mark an error if necessary. No predictions needed to be made, since the participants were able to directly compare letters in the to-be-edited string with the correct letters in the diagram.

The participants in the diagram-animated predictive condition also saw the letters appear one by one in the diagram. However, they were forced to make predictions about which letter would appear next, because after a letter had appeared in the diagram, the participants were not able to mark the corresponding letter in the to-be-edited string as an error. The letters appeared in the diagram at 500-msec intervals. Thus, in the predictive condition, a decision was required before the letter appeared.

The remaining four groups performed the same string-edit task. In the chunk conditions, the letters from the exemplar appeared in the center of the screen, from left to right, in chunks of two or three letters, rather than in a diagram of the grammar. Space was left between the chunks to make them more salient.

The participants in the chunk-static immediate condition saw the entire exemplar, segmented into chunks, at the beginning of each trial. As in the diagram-static immediate condition, the participants were able to compare the letters in the string they were editing with the letters that appeared in the chunked exemplar. Also, as in all the other conditions, a dot appeared under each letter in the to-be-edited string to alert the participants that their time to mark an error in the letter position was almost over. The participants in the chunk-static predictive condition saw only the to-be-edited string at the beginning of each trial. At the end of the trial, the chunked letter string appeared on the screen, meaning that the participants had to edit the entire string before seeing the correct (chunked) string.

In the chunk-animated immediate condition, the chunks appeared one at a time, from left to right. As the chunks appeared on the screen, the participants were able to compare letters in the to-be-edited string with those in the exemplar and to mark any incorrect letters. After a chunk had appeared, the participants had 500 msec per letter in the chunk to mark an error in the corresponding letters in the to-be-edited string.

Finally, in the chunk-animated predictive condition, the chunks appeared one at a time. If the first chunk in the exemplar was three letters in length, the timing dot would move through the first three letters in order, stopping for 500 msec at each letter. When the dot moved to the fourth letter, the first chunk would appear. Thus, the participants were forced to predict what letters would appear in each chunk before seeing the correct chunk of letters appear.

Cued generation test. The cued generation test was identical to that used in Experiment 1.

Grammaticality judgment test. After the cued generation test in the fifth session, the participants completed a grammaticality judgment test. The participants were instructed that the letter strings they

Figure 4. Example states from the chunk-animated predictive condition in Experiment 2.

had seen during the past four sessions followed a set of rules (Reber, 1969). They were told that they would see letter strings on the computer screen and that they should press one key if the letter string followed those rules and another key if the letter string did not follow the rules. The grammaticality judgment test consisted of 100 valid exemplars and 140 invalid lures, which could be divided into two groups. One type of lure was created by substituting a chunk that would be valid in another position (chunk position error). The second type of lure was created by changing one or more of the letters in a chunk to make it invalid (within-chunk error). Eighty of the lures involved chunk position errors, and 60 involved within-chunk errors.

Results and Discussion

As in Experiment 1, we will present the test data from the final session for simplicity. However, we also analyzed the data by session. As will be discussed below, the participants in the diagram-animated conditions generated more perfect strings on the first attempt than did those in the other conditions. The analysis showed that this advantage began early and that group differences were significant beginning at the third session (see Figure 5). With regard to speed, the participants in all the groups generated strings more quickly as training progressed, but there was no significant session $\times$ group interaction. For each dependent measure, we conducted a $2 \times 2 \times 2$ between-subjects ANOVA with display content (diagram vs. chunk), display type (animated vs. static), and prediction (predictive vs. immediate) as factors. Following each factorial ANOVA,

we also conducted a Dunnett's test to determine whether performance in the experimental groups significantly differed from that in the no-training control group.

Cued Generation Test

Means for each of the dependent measures are provided in Table 2.

Acceptable strings. As in Experiment 1, we first looked at generation accuracy, using a moderate criterion. Acceptable strings were measured as the number of letter strings generated on the first attempt that matched at least 70% of the letters in the target exemplar, over the total number of first attempts. Our analysis revealed a trend toward greater generation of acceptable strings in the chunk conditions ($M = .51$) than in the diagram conditions ($M = .43$) [$F(1,158) = 3.66$, $MS_e = .08$, $p < .06$, $\eta_p^2 = .02$]. No other main effects or interactions were significant (all $F_s < 1.5$). A Dunnett's test revealed that all the training groups produced more acceptable strings than did the no-training control. Thus, the data contain some suggestive evidence that the chunk training improved some aspects of performance. We will address this issue later in the context of the grammaticality test results.

Perfect strings. As in Experiment 1, we were most interested in whether training affected the ability to produce strings that met a very high criterion for accuracy. Again, perfects were measured as the number of letter

Figure 5. Production of perfect (100% accurate) strings on the first attempt, by session, in Experiment 2. CAP, chunk-animated predictive; CAI, chunk-animated immediate; CSP, chunk-static predictive; CSI, chunk-static immediate; DAP, diagram-animated predictive; DAI, diagram-animated immediate; DSP, diagram-static predictive; DSI, diagram-static immediate; CNTRL, no-training control.

strings generated on the first attempt that matched 100% of the letters in the target exemplar, over the total number of first attempts. Analyses revealed a significant main effect of display content [$F(1,158) = 5.22$, $MS_e = 0.13$, $\eta_p^2 = .03$]. The participants who saw the diagram at training produced a greater number of perfect exemplars on the first attempt ($M = .14$) than did the participants who saw chunks at training ($M = .08$). The ANOVA also revealed the predicted significant interaction between display type and content [$F(1,158) = 5.0$, $MS_e = 0.12$, $\eta_p^2 = .03$]. Tests of simple effects revealed that the participants who viewed the animated diagram generated more perfect strings per minute ($M = .18$) than did those who viewed the animated chunk display ($M = .07$) [$F(1,77) = 6.5$, $MS_e = 4.99$, $\eta_p^2 = .08$]. When the display was static, the participants who viewed the diagram ($M = .10$) did not generate significantly more perfect strings than did the participants who viewed chunks ($M = .09$) ($F < 1$). No other main effects or interactions were significant. A Dunnett's test showed that only the diagram-animated immediate and diagram-animated predictive groups produced more perfect strings than did the no-training control group. Because these differences between conditions seem to have

emerged early and to have remained consistent across sessions (see Figure 5), there was no evidence that chunking groups were gradually acquiring the higher level knowledge needed for highly accurate performance. Thus, we again replicated the high levels of accuracy observed in Experiment 1 when the participants had the assistance of an animated diagram during training.

The foregoing results present a curious pattern in which chunk training appeared to enhance the production of acceptable strings and (animated) diagram training enhanced the production of perfect strings. This is seen most strikingly in the diagram-animated predictive group, which generated a high number of perfect strings and a relatively low number of acceptable strings. This can be explained by at least two accounts. First, the participants in this condition may have developed some knowledge that was highly accurate and some inaccurate knowledge. Strings generated using the highly accurate knowledge would likely be perfect, whereas those generated using erroneous knowledge would not meet the (more liberal) *acceptable* criteria on the first attempt. The second possibility is that this condition may have facilitated learning in some participants and not in others.

To examine the second hypothesis, we examined individual differences in the number of perfect strings generated on the first attempt. Of the 21 participants in the diagram-animated predictive group, 9 generated no perfect strings on a first attempt. At the other extreme, 7 participants generated a perfect string on more than 25% of their first attempts, and 5 of those participants produced a perfect string on more than 50% of their first attempts. Also, those participants who generated no perfect strings on a first attempt displayed a similarly low level of performance on acceptable strings, generating, on average, an acceptable string on fewer than 4% of their first attempts. In contrast, the high performing participants generated an acceptable string on approximately 83% of their first attempts. Thus, it appears that the diagram-animated predictive training benefited some participants more than others. The participants who generated more perfect strings were likely using their rich knowledge of the grammar to generate highly accurate strings, whereas others fell back on some other, less effective strategy. We will return to this issue in the General Discussion section.

Speed. Finally, we examined the total number of attempts made by the participants per minute during the 20-min cued generation test. Analyses revealed a trend toward a main effect of content [$F(1,158) = 3.75$, $MS_e = 28.48$, $p = .055$, $\eta_p^2 = .02$], since the participants who

Table 2
Performance on the Cued Generation Test in Experiment 2

Performance on the Cued Generation Test in Experiment 2																		
Measure	Chunk								Diagram								Control	
	Animated				Static				Animated				Static					
	Predictive		Immediate		Predictive		Immediate		Predictive		Immediate		Predictive		Immediate			
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD		
Accuracy-acceptable	.50	.26	.52	.26	.55	.23	.48	.22	.42	.37	.50	.31	.41	.32	.38	.28	.08	.12
Accuracy-perfect	.08	.07	.06	.06	.11	.12	.08	.08	.20	.27	.16	.19	.11	.19	.08	.14	.003	.01
Speed	7.92	0.63	7.69	0.61	6.94	0.58	8.23	0.58	9.37	0.60	7.81	0.58	8.20	0.63	8.80	0.60	9.20	0.56

viewed the diagram in training made more attempts at test ($M = 8.56$) than did those who viewed chunks ($M = 7.71$). The ANOVA also revealed a significant interaction between prediction and display type [$F(1,158) = 4.85$, $MS_e = 36.87$, $\eta_p^2 = .03$]. Follow-up tests of simple effects revealed that the participants who had to make predictions while viewing an animated display made more attempts ($M = 8.7$) than did the participants who did so while viewing the static display ($M = 7.6$); however, this difference did not reach conventional significance [$F(1,82) = 3.3$, $MS_e = 7.9$, $p = .073$, $\eta_p^2 = .04$]. The participants who did not need to make predictions did not significantly differ on number of attempts, regardless of display type [$M = 7.7$ and 8.5 for the animated and the static conditions, respectively; $F(1,82) = 1.8$, $MS_e = 7.4$, n.s., $\eta_p^2 = .02$]. No other main effects or interactions reached significance. A Dunnett's test revealed that the experimental groups did not differ from the no-training control group. Overall, the results suggest that the groups were highly similar in terms of speed. Thus, the diagram-animated conditions, which displayed high levels of accuracy, were just as fast as comparable conditions.

Grammaticality-Judgment Test

We also examined grammar knowledge using grammaticality judgments (see Table 3). Overall grammaticality judgment accuracy was measured as the proportion of letter strings correctly classified as valid or invalid. Analyses revealed a main effect of content [$F(1,158) = 4.42$, $MS_e = .03$, $\eta_p^2 = .03$]. The participants who saw chunks in training ($M = .67$) were more accurate than those who saw the diagram at training ($M = .63$). No other main effects or interactions were significant. A Dunnett's test showed that all groups except that in the diagram-static immediate condition responded more accurately than did the no-training control group.

We also examined performance on the grammaticality-judgment test in a more fine-grained way by looking at performance on strings that were invalid in specific ways. First, we examined identification accuracy for strings that were invalid because a (valid) chunk was in the wrong position. Analyses revealed a significant main effect of content [$F(1,158) = 10.89$, $MS_e = .13$, $\eta_p^2 = .07$]. The participants who saw a diagram at training ($M = .58$) were more accurate than participants who saw the exemplars parsed into chunks ($M = .52$). There was also a significant main effect of prediction [$F(1,158) = 4.34$, $MS_e = .05$, $\eta_p^2 = .03$], since

the participants who made predictions at training ($M = .57$) were more accurate than the participants who did not make predictions ($M = .53$). Although there were no significant interactions, there was a trend toward an interaction between content and display [$F(1,158) = 3.32$, $MS_e = .04$, $p = .07$, $\eta_p^2 = .02$]. Given that this interaction was critical to our hypotheses, we examined the simple effects. As is obvious from Table 3, the participants who had the assistance of a diagram that was animated identified significantly more chunk position errors ($M = .60$) than did the participants who were assisted by animated chunks ($M = .51$) [$F(1,77) = 11.66$, $MS_e = .01$, $\eta_p^2 = .13$]. However, the participants who had static versions of the diagram ($M = .55$) or the chunks ($M = .52$) did not differ in identification accuracy ($F = 1.3$). A Dunnett's test showed that only the diagram-animated predictive group responded more accurately on strings with chunk position errors than did the no-training control. Thus, the results suggest that diagram training increased higher level knowledge of the grammar, relative to chunk training.

We next examined performance when there was an error within a chunk in the string. In other words, these were invalid exemplars that contained a letter that could not appear in a given position according to the rules of the grammar. In this analysis, a different pattern of results emerged. There was a significant main effect of content [$F(1,158) = 10.24$, $MS_e = .24$, $\eta_p^2 = .06$]. In this case, the participants who saw the exemplars parsed into chunks at training ($M = .75$) were more accurate than the participants who saw a diagram ($M = .67$). There were no other significant main effects or interactions. A Dunnett's test revealed that all the groups that saw chunks during training performed more accurately than the control group, whereas those that saw the diagram at training did not significantly differ from control. Thus, the results suggest that chunk training increased knowledge of the dependencies within chunks, relative to the participants who were trained using the diagram.

Relationship between grammaticality-judgment test and cued generation test performance. The results of each of the two tests of grammar knowledge revealed differences as a function of type of training. Training assisted by an animated diagram led to higher generation of perfect strings on the cued generation test and more accurate identification of invalid strings with chunk position errors. Chunk training appeared to benefit the production of acceptable strings on the cued generation test and led

Table 3
Performance on the Grammaticality-Judgment Test in Experiment 2

Measure	Performance on the Grammaticality Judgment Test in Experiment 2																	
	Chunk								Diagram								Control	
	Animated				Static				Animated				Static					
	Predictive		Immediate		Predictive		Immediate		Predictive		Immediate		Predictive		Immediate			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Overall accuracy	.69	.07	.66	.10	.66	.07	.66	.07	.67	.15	.63	.11	.63	.10	.60	.12	.54	.05
Accuracy—chunk position errors	.53	.03	.50	.02	.55	.02	.50	.02	.63	.02	.58	.03	.56	.02	.54	.02	.53	.02
Accuracy—within-chunk errors	.80	.03	.73	.03	.75	.03	.72	.03	.69	.03	.69	.03	.68	.03	.64	.03	.59	.03

Note—Eighty items involved chunk position errors, and 60 involved within-chunk errors.

to more accurate identification of invalid strings with errors within a chunk on the grammaticality test. Thus, the pattern of results suggests that these respective findings might be related. In particular, we hypothesized that good within-chunk knowledge would facilitate the production of acceptable strings on the cued generation test because participants could use such knowledge to generate adjacent letters to the provided letter cues. However, such knowledge would typically not be enough to produce perfect strings. Instead, higher level knowledge would be needed to produce such highly accurate strings on the first attempt.

In order to explore this question, we examined partial correlations between the two tests collapsed across conditions.² Analyses revealed a significant and strong correlation between acceptable string production and identification accuracy for within-chunk errors [$r(163) = .48$] when controlling for identification accuracy on chunk position errors. However, the complementary correlation between acceptable string production and identification accuracy on chunk position errors was not significant [$r(163) = .02$]. A similar analysis on perfect string production revealed a significant correlation with identification accuracy for chunk position errors [$r(163) = .40$] when controlling for identification accuracy on within-chunk errors. The complementary correlation between perfect string production and identification accuracy on within-chunk errors was substantially smaller and was not significant [$r(163) = .14$]. These analyses suggest that one reason why the animated-diagram participants were better able to produce perfect strings on a first attempt is that they acquired better higher level grammar knowledge during training. Similarly, one reason why the chunk-training participants appeared better able to produce acceptable strings on a first attempt is that they acquired better knowledge of within-chunk dependencies during training.

GENERAL DISCUSSION

In contrast to prior research (Domangue et al., 2004), two experiments showed that it is possible for learners to build rich and accurate knowledge of a finite-state grammar and to quickly and accurately generate exemplars. Furthermore, our results demonstrate the importance of the level of knowledge emphasized (letter, chunk, or structural information) during training and the manner in which such information is made available to the learner. In Experiment 1, the participants who were assisted by an animated diagram while editing training strings (diagram assistance) were more likely to produce completely accurate strings on their first attempt during a cued generation test than were the participants whose letter-level responses were primed during training (letter assistance) or whose responses were not primed (no assistance) but were just as fast at producing exemplars at test. In Experiment 2, the participants who received the animated diagram produced more completely accurate strings than did the participants who received a static version of the diagram or those who were aided by prechunked strings. Furthermore, the results of the grammaticality test suggest that the animated diagram led to enhanced higher level structural knowledge

of the grammar, since the participants were significantly more accurate at identifying chunk position errors. Thus, the participants achieved both fast and accurate string generation only if they were aided by structural (diagram) information during training and relevant information was highlighted just when it was needed (using animation). We next will discuss the implications of these findings for a number of important issues.

As was noted in the introduction, our manipulations of training conditions were motivated by current theories of artificial grammar learning and by an analysis of previous failed attempts to produce both fast and accurate performance (Domangue et al., 2004). Our results were most supportive of the notion that a combination of top-down (model-based) and bottom-up (experience-based) processing might be a particularly powerful way to increase the acquisition of expertise (e.g., Sun, 2002; Sun et al., 2005). However, our work also suggests that integrating these types of knowledge is not easy. It is clear that simply providing information about the underlying structure of the grammar and exposing participants to exemplars during training is not sufficient. In Domangue et al. (2004), participants who entered exemplars into a grammar diagram ended up with moderate levels of speed and accuracy. Furthermore, participants from our study who received static versions of the diagram did not achieve the high levels of accuracy obtained by those who received the animated versions, nor did they demonstrate good knowledge of the higher level structure of the grammar on the grammaticality test (Experiment 2). Thus, it clearly matters how model-based (structural) information is provided during training.

Comparing performance in the present experiments and in Domangue et al. (2004) provides some clues regarding the factors that underlie the effectiveness of the animated diagram during training. We emphasize two main features that, together, play a role in its utility. First, the participants in our animated diagram conditions practiced using the diagram to help them edit the training strings. In contrast, participants in Domangue et al.'s GR condition memorized the whole diagram and then performed the cued generation task by mentally recalling and working through the diagram to generate strings (which resulted in accurate but very slow performance).

Thus, one key to the achievement of fast and accurate performance involves using the model to acquire procedural knowledge of how to generate strings directly (without needing an explicit representation of the model), rather than learning to rely on an explicit representation of the model to generate strings. However, our results suggest that this characteristic alone is insufficient. A second important factor is that the animated diagram provides structural knowledge about the grammar in a just-in-time fashion. The present set of experiments emphasized speed during training. In order to complete the string-edit task, the participants were required to make decisions quickly about the accuracy of a given letter within the presented strings. We believe that it is highly likely that the restricted time led the participants to encode those aspects of the diagram that could be quickly and easily used to perform the string-edit task. Furthermore, performance was en-

hanced in the animated-diagram conditions, but not in the static conditions. The difference between the conditions is that, in the animated versions, the relevant part of the diagram was highlighted just as the corresponding part of the training string was to be edited. Thus, we suggest that it was the combined emphasis on using the diagram to assist performance in a fast-paced training task and the focus on the relevant part of the diagram provided by animation that allowed the participants to achieve fast and accurate performance.

We note that although the animated-diagram groups, as a whole, showed highly accurate performance (and similar levels of production), the results for the diagram-animated predictive group suggest that there may be individual differences. Specifically, some participants in this condition generated very high levels of perfects and acceptable strings, whereas others did quite poorly on both measures. One reasonable candidate explanation for these findings concerns differences in working memory capacity. Relative to those in the diagram-animated immediate condition, the participants in the predictive condition had to maintain information about the upcoming letter in the target string while attempting to use the diagram to narrow down the potential candidate letters. Given the speed of the editing task, it is likely that participants with lower working memory spans would have been at a disadvantage and, so, may have relied instead on less effortful (but less optimal) strategies to do the training task. A more definitive answer to this question awaits future research.

Some models of artificial grammar learning (e.g., Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990) suggest that grammar knowledge is acquired through bottom-up learning from exposure to exemplars. Under this conception, higher level grammar knowledge (needed to support highly accurate performance) is likely to be acquired with extensive practice and through enhanced processing of lower level knowledge, which allows for higher level knowledge to be built. We attempted such facilitation by enhancing the processing of lower level (letters; letter assistance condition, Experiment 1) and mid-level (chunks; Experiment 2) grammar knowledge. Although both experiments involved extensive practice (considerably more than most studies of artificial grammar), we did not obtain fast and accurate performance. In Experiment 2, the chunking manipulation improved lower level grammar knowledge, as indicated by the participants' performance on intrachunk errors. However, the participants did not appear to "build" upon this knowledge to increase their knowledge of higher order aspects of the grammar. Although we leave open the possibility that enhanced bottom-up processing would eventually increase higher order knowledge given enough time, this mode appears to be less efficient (within the constraints of this study) in producing fast and accurate performance than is providing structural knowledge during training.

Our findings *are* consistent with the notion that participants in typical artificial grammar studies learn about the grammar through the process of parsing exemplar strings into chunks (e.g., Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). In Experiment 2, training

that emphasized chunking led to good performance on the grammaticality test, and the participants in these conditions were particularly accurate at detecting within-chunk errors. Furthermore, chunk training led to fast and moderately accurate performance on the cued generation test. However, our results also suggest that knowledge of the composition of chunks is not sufficient for highly accurate string generation. Relative to the diagram-training conditions, the participants in the chunk-training conditions were poorer at identifying strings that had chunk position errors. Even the participants who received animated chunk training did not achieve the high level of accuracy achieved in animated diagram conditions on the cued generation test. Thus, although learning chunks and their dependencies can lead participants to achieve moderate knowledge of the grammar, higher levels of string generation accuracy were possible only when the participants acquired some understanding of its underlying rules by means of the diagram.

The big picture emerging from our pattern of results suggests that bottom-up learning has a downward bias. In Experiment 1 when the task focused the participants' attention on letter-by-letter associations (the letter assistance condition), they did not acquire the higher level knowledge required to produce high numbers of perfect strings. In Experiment 2, when the task focused attention on chunks by presenting strings "prechunked," the participants learned a great deal about intrachunk structure but less about higher level knowledge (e.g., chunk positions). Only when the animated diagram was used to enhance learning was knowledge of chunk positions well learned. We do not deny that true bottom-up learning (i.e., in an upward direction) can occur implicitly; this has been well demonstrated, for example, in serial pattern learning, where participants learn about regularities in sequences of stimuli over trials (Stadler, 1993). However, our data suggest that the experience-based learning system is more efficient for learning about structure within units, rather than between nominal stimuli (e.g., chunk position). Thus, our results suggest that adding model-based knowledge is most likely to enhance learning when it can inform the learner about higher level structural information that is more difficult to acquire implicitly (through experience-based learning).

Although there has been prior research claiming facilitative effects of animation on learning (e.g., Large, Beheshti, Breuleux, & Renaud, 1996; Schnotz, Böckheler, & Grzondziel, 1999), it has been argued that this research does not demonstrate that animation improves learning because of methodological confounds (Tversky, Morrison, & Betrancourt, 2002). For example, some studies present additional information in animated versions than in static versions of the same stimuli (e.g., Large et al., 1996) or allow interactivity in animated versions, but not in static versions. When these confounds are not present, the typical finding is that animation is not beneficial (e.g., Rieber, Boyce, & Assad, 1990). In our study, participants within a given content condition (chunk or diagram) received the same information in the animated and static versions of the display, and the level of interactivity (i.e., what could

be done with the interface) was the same across all the conditions. The results of this experiment demonstrate a facilitative effect of animation in the diagram conditions, but not in the chunk conditions. These findings suggest that animation may not be of help in situations in which the target information can already be acquired in a simple direct fashion, as in our chunk conditions. However, our findings suggest that animation may be helpful when the information would not be normally acquired through experience, when the target information is complex (e.g., when multiple sources of information must be integrated) and the animation focuses the attention of the learner on aspects of the information that are most relevant for that particular point in time.

Finally, our results have practical implications. For example, we found that providing structural knowledge about a complex grammar “just in time” during training leads to accurate and quickly accessed representations but that providing the same information ahead of time (Domangue et al., 2004) or asking participants to attempt to discover the underlying rules (Mathews et al., 1989) is much less effective. Although our artificial grammar may seem different from topics normally taught in the classroom, such findings appear consistent with research suggesting that even topics whose underlying structure is relatively explicit and salient are typically learned better with guided instruction than with relatively unguided discovery learning (Klahr & Nigam, 2004; for a review, see Mayer, 2004). Furthermore, given the increased use of labs and other experiential learning techniques for instruction, the issue of how model- and experience-based processes interact becomes an extremely important one. The more we understand the mechanisms that facilitate learning under such conditions, the more likely it is that we can achieve the goal of producing expert performance on a consistent basis.

AUTHOR NOTE

This research was partially supported by Army Research Institute Grant DASW01-00-K-0012 to R.S. and R.C.M. We thank Emily Elliott, Janet McDonald, and Melissa Beck for their insightful comments on an earlier version of the manuscript. Correspondence concerning this article should be addressed to B. Sallas, R. C. Mathews, or S. M. Lane, Department of Psychology, Office of Applied Cognition, Louisiana State University, 236 Audubon Hall, Baton Rouge, LA 70803 (e-mail: sallas@lsu.edu, psmath@lsu.edu, slane@lsu.edu).

REFERENCES

- ANDERSON, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, **89**, 369-406.
- ANDERSON, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- BERRY, D. C., & DIENES, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hillsdale, NJ: Erlbaum.
- DOMANGUE, T. J., MATHEWS, R. C., SUN, R., ROUSSEL, L. G., & GUIDRY, C. E. (2004). Effects of model-based and memory-based processing on speed and accuracy of grammar string generation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 1002-1011.
- ELLIS, N. C., & LAPORTE, N. (1997). Contexts of acquisition: Effects of formal instruction and naturalistic exposure on second language acquisition. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 53-83). Mahwah, NJ: Erlbaum.
- ERICSSON, K. A., KRAMPE, R. T., & TESCH-RÖMER, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, **100**, 363-406.
- FRENSCH, P. A., HAIDER, H., RÜNGER, D., NEUGEBAUER, U., VOIGT, S., & WERG, J. (2003). The route from implicit learning to verbal expression of what has been learned: Verbal report of incidentally experienced environmental regularity. In L. Jiménez (Ed.), *Attention and implicit learning* (pp. 335-366). New York: Benjamins.
- KLAHR, D., & NIGAM, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, **15**, 661-667.
- KLEIN, G. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.
- 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, & Cognition*, **22**, 169-181.
- LANE, S. M., MATHEWS, R. C., ROUSSEL, L. G., SALLAS, B., PRATTINI, R., & SUN, R. (in press). Facilitative interactions of model and experience-based processes: Implications for type and flexibility of representations. *Memory & Cognition*.
- LARGE, A., BEHESHTI, J., BREULEUX, A., & RENAUD, A. (1996). The effect of animation in enhancing descriptive and procedural texts in a multimedia learning environment. *Journal of the American Society for Information Science*, **47**, 437-448.
- LEWICKI, P., CZYZEWSKA, M., & HOFFMAN, H. (1987). Unconscious acquisition of complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 523-530.
- 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, & Cognition*, **15**, 1083-1100.
- MATHEWS, R. C., & COCHRAN, B. P. (1998). Project grammarama revisited: Generativity of implicitly acquired knowledge. In M. A. Stadler & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 223-259). Thousand Oaks, CA: Sage.
- MAYER, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, **59**, 14-19.
- MILLER, G. A. (1958). Free recall of redundant strings of letters. *Journal of Experimental Psychology*, **56**, 485-491.
- PERRUCHET, P., & PACTEAU, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, **119**, 264-275.
- PERRUCHET, P., VINTER, A., PACTEAU, C., & GALLEGO, J. (2002). The formation of structurally relevant units in artificial grammar learning. *Quarterly Journal of Experimental Psychology*, **55A**, 485-503.
- REBER, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior*, **6**, 855-863.
- REBER, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, **81**, 115-119.
- REBER, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.
- RIEBER, L. P., BOYCE, M. J., & ASSAD, C. (1990). The effects of computer animation on adult learning and retrieval tasks. *Journal of Computer-Based Instruction*, **17**, 46-52.
- SCHNOTZ, W., BÖCKHELER, J., & GRZONDZIEL, H. (1999). Individual and co-operative learning with interactive animated pictures. *European Journal of Psychology of Education*, **14**, 245-265.
- SEGER, C. A. (1994). Implicit learning. *Psychological Bulletin*, **115**, 163-196.
- SERVAN-SCHREIBER, E., & ANDERSON, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 592-608.
- SHANKS, D. R., & ST. JOHN, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral & Brain Sciences*, **17**, 367-447.
- SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.

- STADLER, M. A. (1993). Implicit serial learning: Questions inspired by Hebb (1961). *Memory & Cognition*, **21**, 819-827.
- SUN, R. (2002). *Duality of the mind: A bottom up approach toward cognition*. Mahwah, NJ: Erlbaum.
- SUN, R., SLUSARZ, P., & TERRY, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, **112**, 159-192.
- TVERSKY, B., MORRISON, J. B., & BETRANCOURT, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, **57**, 247-262.
- WILLINGHAM, D. B., NISSEN, M. J., & BULLEMER, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 1047-1060.

NOTES

1. We have adopted the terms *experience-* and *model-based* processing (Lane et al., in press) to capture the results of research that suggest that complex mental skills are learned and deployed through the use of two different and complementary types of processes (and resulting representations; see, e.g., Anderson, 1982; Berry & Dienes, 1993; Mathews et al., 1989; Reber, 1993; Sun, 2002; for arguments against, see, e.g., Shanks & St. John, 1994). We have adopted these terms (as opposed to other terms, such as implicit/explicit; see Reber, 1993) for two major reasons. First, although experiential knowledge is largely nonconscious, people can and often do access this information by observing the cues they are attending to while performing a task (see Frensch et al., 2003, for a related but somewhat different approach). Second, although we argue that these two categories are psychologically meaningful, we also

assume that both types of processes operate in nearly all tasks (e.g., Mathews, 1997; see also Reber, 1993; Seger, 1994; Willingham, Nissen, & Bullemer, 1989; but see Lewicki, Czyzewska, & Hoffman, 1987) and, thus, the potential interaction of these processes should be considered.

2. We also examined whether the group differences obtained on the cued generation test measures would be impacted when identification accuracy for a given lure type was entered as a covariate in an ANCOVA. Beginning with production of acceptable strings, recall that the results of the factorial ANOVA revealed a trend ($p < .06$) toward greater performance in chunk than in animated training. An ANCOVA with identification accuracy on within-chunk errors revealed no significant main effects or interactions, and the associated F for the main effect of display content was near zero. When a complementary ANCOVA was conducted with identification accuracy on chunk position errors as a covariate, the main effect of display content was significant at $p < .01$ (no other main effect or interactions were significant). We next examined production of perfect strings. The original factorial ANOVA revealed a main effect of display content and a content $\times$ display type interaction. An ANCOVA with identification accuracy on chunk position errors as a covariate revealed no significant main effects or interactions. A complementary ANCOVA with identification accuracy on within-chunk errors as a covariate revealed main effects of display content and a content $\times$ display type interaction (no other effects or interactions were significant). Thus, for each measure of performance on the cued generation test, we observed the same pattern. When we conducted an ANCOVA using the hypothesized related grammar knowledge type as a covariate, the ANOVA results did not obtain. However, when the alternative grammar knowledge type was used, the pattern of results was the same as for the ANOVA.

APPENDIX

List of Chunks Used in Experiment 2

CVC	P(T)	VV
CXP	(S)	XX
CXT	(T)	XS
SCP	TSX	TXX
SCT	VPS	P
T(S)	(VPX)	

Note—Parentheses denote a letter of the chunk that can be repeated.

(Manuscript received October 7, 2006;
revision accepted for publication February 21, 2007.)