ELSEVIER

Journal of Cognitive Systems Research 1 (2000) 161–174

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

The role of implicit learning in the acquisition of generative knowledge

Action editor: Vasant Honavar

Robert C. Mathews*, Lewis G. Roussel, Barbara P. Cochran, Ann E. Cook, Deborah L. Dunaway

Department of Psychology, Louisiana State University, Baton Rouge LA 70803-5501, USA

Received 30 July 1999; accepted 20 August 1999

Abstract

This paper reports two experiments using a modified artificial grammar paradigm to study the acquisition of generative knowledge. It employs a starship game and a computerized form of motherese to study acquisition of generative knowledge of a grammar. It was found that initial exposure to a small set of representative cases and an opportunity to organize the cases enhanced ability to generate grammatical strings. Neither seeing a large number of cases nor building a conceptual model of the grammar enhanced acquisition of generative knowledge. However, building a conceptual model did increase learners' confidence in their ability to generate strings. The quality of strings generated by learners remained fairly constant across attempts, suggesting that their performance reflects true generative knowledge rather than memories of exemplars.

1. Introduction

We are interested in the processes that lead to acquisition of knowledge that can be applied fluently and used adaptively within a domain, what we call generative knowledge. This is the type of knowledge that distinguishes experts from novices, chefs from cooks. Generativity is defined by Corballis (1991) as the ability to produce an unlimited number of representations from a small number of components and a set of rules for combination of components. Common sense would suggest that generative knowledge depends on conscious, explicit, or symbolic knowledge of a domain. Thus, we are much more creative than chimpanzees, for example, because we have language and symbolic thought to enhance our creativity. However, research on experts, such as chess masters, also suggests that non-conscious, implicit, knowledge is also important. For example, chess masters often 'see' good moves rather than consciously figuring them out or mentally testing possible strategies (Degroot, 1966; Simon & Gilmartin, 1973). Similarly, we acquire language fluency more from experimental practice than conscious reflection about grammatical structure.

The literature on learning without conscious knowledge or control falls under the heading of implicit learning research. Most of this research supports the notion that implicitly acquired knowl-

*Corresponding author.

E-mail address: psmath@unix1.sncc.lsu.edu (R.C. Mathews)

1389-0417/00/$ – see front matter
PII: S1389-0417(00)00007-3

edge is relatively inflexible, making it a poor candidate for supporting generative use of knowledge. However, we believe this alleged rigidity of implicit knowledge is partly an artifact of the way it has been studied. Searching for pure cases of implicit learning (situations where there is no evidence of conscious awareness of what was learned), has moved this research paradigm toward studying relatively simple motor tasks (e.g., reaction time in serial pattern learning tasks, e.g., Cleeremans & Jimenez, 1998; Stadler, 1995). These simple tasks do not lend themselves to studying generativity. In real life, implicit knowledge gained from experience does not come in pure implicit or explicit modes. Instead, our experiential knowledge (much of which is implicitly acquired) is mixed with our conceptual or reflective knowledge of the domain.

Even though these knowledge types occur in combinations, they remain somewhat independent because they are adapted to different functions. Experiential knowledge is most adapted to guide behavior in complex tasks, taking in rich environmental cues at a rapid rate and instantly guiding behavior. Reflective or symbolic knowledge evolved for communication and preservation of knowledge. Obviously, there must be some coordination between these two knowledge systems. Yet, the two types of knowledge are not always consistent with each other, even in the minds of the best experts. When conflicts exist, it is usually the richer experiential knowledge that is used to perform tasks (e.g., Stanley, Mathews, Buss, and Kotler-Cope, 1989) while the more abstract explicit knowledge is used to communicate about the domain (Boshuizen & Schmidt, 1992). Consequently, experts often give incomplete or inaccurate descriptions of their own methods, creating the familiar knowledge engineering bottleneck (Nikolopoulos, 1997). If largely implicit, experiential knowledge is what dominates human performance on complex tasks, (for example a jazz musician or a skilled potter) then implicit knowledge must not be as rigid as researchers believe and it may play a crucial role in development of generative knowledge.

Acquisition of these two types of knowledge seems to occur at different rates and they depend on different variables. Acquiring experiential knowledge depends most on exposure to good examples or cases. Thus, seeing experts work on problems, working on problems as an apprentice, or tackling problems on your own contribute most to experiential knowledge. Reflective knowledge comes from others (books or instructions) and from our own reflective thoughts as we attempt to consciously think about concepts in the domain. We believe that development of high levels of generative knowledge requires both types of knowledge working together synergistically as in hybrid learning systems (Sun, 1995).

In many situations when learners tackle a new complex domain, implicit experiential knowledge precedes explicit knowledge (we learn more than we can tell, e.g., Mathews, Buss, Chin & Stanley, 1988). We think this occurs because implicit knowledge is related to pattern recognition and perception (see Mathews & Cochran, 1998). Swinging from tree to tree, there is no time for slow cognition (e.g., logical analysis of situation). Implicit thinking evolved to occur on the fly in order to succeed in our fast paced environment. Since the environment is variable, generativity must be built into the system. Through implicit learning a person acquires a sense of what varies, what components of tasks remain stable, and how task components can be effectively interrelated. We begin to literally 'see' problems in the domain as sets of components that can be recombined flexibly under certain conditions.

In order to demonstrate the importance of implicit learning in generative knowledge we need a laboratory paradigm that is rich enough to compare generative abilities acquired by seeing examples (e.g., examining and memorizing cases) versus trying to develop a conceptual model of the task (reflective knowledge). The measure of performance must also go beyond mere memory or generalization from memorized cases. Demonstrating that acquired knowledge is really generative rather than merely using memories of past cases can be very difficult (see Vokey & Brooks, 1992).

George Miller (1968) invented an interesting paradigm for studying the development of generative grammatical knowledge. This paradigm involved having subjects attempt to generate sequences of letters that match those that can be generated by a finite state grammar, such as the one shown in Fig. 1. After each attempt participants got feedback about whether the string they generated was valid or

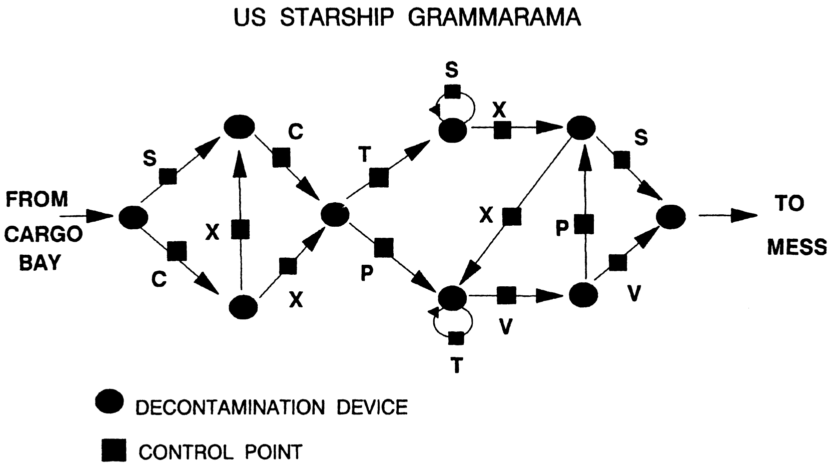

Fig. 1. Transition map of the artificial grammar used in the Poison Food Game

invalid. However, in Miller's experiments it quickly became obvious that only the very simplest grammars could be mastered in the short time period that one could get subjects to work on this rather boring task It was also very frustrating to generate many strings and be constantly told ''wrong, wrong, wrong...''. Yet subjects were only given dichotomous feedback — right or wrong — patterned after the procedure used in the concept formation literature. Thus the experiments consisted of mostly failed attempts to generate valid strings. Quoting Miller (1968), ''Moreover, a person who is told twenty or thirty times in a row that he is wrong has an understandable urge to strangle the experimenter'' (p. 144).

Early on, Miller realized that some type of automation was essential for effective grammar production experiments. He enlisted the help of E.B. Newman to design what must have been one of the first precursors of the laboratory computer that runs most experiments today. Newman's black box enabled subjects to type in strings and get feedback (right or wrong) after each response. However, being on the edge of technology at the time, it didn't work reliably enough to make the experiments practical. Later when the first reliable laboratory computers did become available, Miller again tried to create an effective paradigm for studying grammar production. He once again patterned the design of his experiments after the concept learning tradition. His subjects generated strings and they received dichotomous feedback. To avoid long training periods, the grammars were kept simple. In retrospect both of these approaches (dichotomous feedback and oversimplified grammars) probably contributed to the lack of interesting results.

Once again, Miller became frustrated with the task. He wrote: ''Surely we have here discovered the most inefficient way to teach a set of rules — the way of pure induction — almost beautiful in its unadorned ugliness. Those poor subjects, working so long, thinking so hard, failing so often... But even beyond pragmatic questions of efficiency it is difficult to suppress a feeling that the whole interaction is grotesquely stupid'' (Miller, 1968; p. 163). In retrospect, part of his participants' frustration may have involved thinking too much — relying too much on reflective versus experiential learning.

Various ways of circumventing these problems were tried in subsequent studies. For example, Eva Shipstone avoided giving negative feedback entirely by having subjects devise their own classification schemes for sets of valid strings printed on cards. She also had them talk aloud, generating rich protocols on their developing classification schemes. However, rich as these protocols were in detail, they seemed to offer little in generalizations across subjects. Mostly it seemed that idiosyncratic strategies were being employed. Another approach, begun by Reber (1967, 1969, 1976, 1993) laid the foundation for the extensive research on implicit learning that followed. Reber made the learner's task easier by showing them sets of valid strings rather than trying to learn by self generated strings. He also switched the test to an easier string discrimination test (like a multiple choice test), rather than string generation. While this literature is rich and useful, the knowledge acquired in this paradigm doesn't require the level of generativity of Miller's original task (e.g., having subjects actually generate strings). Therefore, we set out to solve the problems with the original Miller paradigm and use it as a tool to study acquisition of generative knowledge.

2. Redesigning the Artificial Grammar Paradigm: The Poison Food Game

One major obstacle to studying generativity with artificial grammars was the absence of a meaning element. The lack of meaning of generated strings is related to a second recurrent problem — the boring nature of the artificial grammar learning task. We attempted to solve both of these problems by incorporating generation of strings into a game that gave meaning to the strings and made the task more exciting. Our cover story places the task in a starship. Subjects were given instructions containing the following story:

> *We are on a military transport vessel attempting to bring remnants of a space colony back home. Unfortunately, we are short on food for the long trip. Making matters worse, much of the food that we took on board from the colony has been contaminated by a radioactive poison. Your job is to learn to distinguish poison from non-poisoned food by recognizing poison food labels. Since one of our goals was to contrast experiential learning with reflective learning, we developed two versions of the poison food game: Explicit groups (reflective learning) were told that the poisoning of the food was done by spies and that spies have a secret code for discriminating poison from non-poison food labels. We know this because spies never eat poison food. Thus, explicit groups were encouraged to reflect and crack the spies' code to discriminate poison from non-poison food labels. In the implicit groups (experiential learning) participants were led to believe that the food poisoning was accidental, there was no mention of spies. They were also led to expect that memorizing poison can labels would be helpful because poison can labels 'look alike'. Thus, these subjects were encouraged to focus their attention on looking at and memorizing examples of poison food labels, emphasizing less reflection about the structure of the codes used to label the cans of food.*

We also attached meaning to the individual letters of strings in terms of movements of the food through control points about the starship. Thus, we added something like a meaningful graphemic level to the strings. This was done by elaborating the starship story as described below:

> *The food taken on board our vessel came originally from another vessel on which all of the passengers died from the poisoned food. Before they all perished, in a last effort to save themselves, members of that ship had installed decontamination devices throughout the ship. These decontamination devices were placed at several control points on the ship where food was moved from one location to the next. However, many of the decontamination devices were inoperative. Every can of food that passed through at least one working decontamination device in its travels about the ship, was and still is safe to eat. Cans that passed through only non-working decontamination devices are still poisonous and must not be eaten.*

The strings of letters represent labels on cans of food that were stamped as the food was moved from one control point to another. The stamps indicate the squad that the food inspector at each control point belonged in. Thus, the same letter could be used by more than one inspector at different control points. Nonetheless, the letter sequences in the labels remain as a cryptic but accurate record of movements of the cans of food around the large ship. Your task is to learn to recognize poison food cans so that we can avoid feeding them to the people on board our vessel.

The second key problem that blocked progress with Miller's paradigm was the fact that it takes too long to learn complex artificial grammars, making it practically impossible to do experiments in the short time we can keep volunteer subjects in the lab. Yet, if one resorts to excessively simple grammars, as Miller did in his original studies, you lose the complexity where implicit processes are needed. A related problem was that continued failure in attempts to generate valid strings, which is bound to occur when subjects attempt to generate strings from a complex grammar without any previous experience, is very frustrating and discouraging. In our attempt to solve these problems, we took an idea from mothers' interactions with their toddlers learning to speak. Mothers often help children by guessing what they are trying to say and repeating the inferred message back to the child. Our computerized version of 'motherese' also did not require a completely accurate generated string for positive feedback. Instead, whenever a subject generated a string that was close to a valid string, we had the computer find the nearest valid string and return the corrected string to the subject. This form of computerized motherese, we thought would facilitate the learning process and avoid uninteresting search behaviors (e.g., trying slight changes in a generated string to find an exact match to a valid string) that dominated the behavior of subjects in the earlier experiments (Miller, 1968). Of course, our motherese idea had to be implemented in the starship game. To accomplish this the following information was included in the instructions:

> *The poisoned food is highly radioactive. Although all of the food supply was initially contaminated, each time it is passed through a working decontamination the amount of radioactivity is reduced. Thus, when tested with a special geiger counter on the ship, radioactivity levels in individual cans of food may range from* 0 *to* 10. *Each can label generated during training will be located by computer and tested for radioactivity. Only cans that test at level* 10 *are poison. Any can with a radioactivity reading lower than* 10 *is safe to eat. Also, since cans that have readings above* 7 *are similar to a poison can label* (*a* 10), *the computer is capable of tracking down the related poison can and giving the exact label.*

The radioactivity reading returned by the computer for each label was ten times the proportion of letters in a generated string that matched a valid string not previously found by the subject. Since valid strings get readings of 10 and the score declines with fewer matching letters, it is also a simple measure of string goodness or degree of grammaticality of the generated string. Once a subject found a poison can of food, the computer eliminated that can from its list of valid strings. Thus, if a subject typed the same can label again, the computer would not find the same poison can again. Instead the reading would be based on the closest remaining valid string not yet generated by the subject.

3. Experiment 1

Many previous studies have examined people's ability to discriminate valid from invalid strings following a training period of either memorizing examples (implicit learning groups) or trying to discover the rules of the grammar (explicit learning groups). These studies have had varied results including finding an interfering effect of looking for rules, finding no effect, or even finding a beneficial effect (see Reber, 1993, p. 47 for an overview of these studies). The general interpretation of these effects is that looking for rules will help if subjects can find them (Reber, 1993). In studies where rule finding did not help, participants were not provided with any notions about the type of rules they should took for, nor are they given the opportunity to use paper and pencil to help figure out the rules. Norman (1993) has noted that testing mental capacities in the absence of the normal cognitive artifacts on which our thinking depends may give false impressions about human abilities.

In this experiment we tried to optimize the benefits of looking for rules. We gave our explicit groups, who we will call model builders, instructions on what types of rules to seek. We showed them how to build a transition map of a grammar from the use of paper and pencil to decipher the rules of the grammar. Also, they were allowed to keep their map of the grammar and use it on the generation test. Our expectation was that this optimal environment for rule learning would allow their reflective knowledge of the grammar to develop in conjunction with their implicit knowledge gained from viewing a large set of strings. We anticipated that these participants would perform much better than the implicit groups who only attempted to memorize the set of study strings.

3.1. Method

Ninety-five subjects participated in the experiment. They were undergraduate students at Louisiana State University who were given extra credit points in return for their participation. Subjects were randomly assigned to a learning mode (implicit or explicit) and to one type of practice test (discriminate or generate). The experiment consisted of four sessions. Each of the first three sessions consisted of a 30 minute study period followed by one of the two types of practice test. The final session consisted of a one hour test. Half of the subjects in each group took a final discrimination test, which was just like the practice discrimination test except that it included all 176 strings generated by the grammar and no feedback was provided. The remaining subjects, those of most interest in this paper, took the final generation test in which they attempted to generate as many of the strings (poison food labels) as they could in one hour.

Subjects were assigned to one of two learning modes: model builders (explicit rule finding) or memorizers (implicit learners of grammatical structure). Model builders were told to develop an unsafe path map during their study periods by combining information from different study exemplars. Fig. 1 provides a complete unsafe path map, listing all the failed decontamination units. Thus, any can of food that moved about the ship only along these paths is still poison. Any can of food that is moved through any path (from one control point to another) that is not included on this map must have gone through at least one safe decontamination unit and it is safe to eat. The map in Fig. 1 is also a representation of the transition matrix of the finite state grammar. This grammar, which was previously used by Mathews et al., (1989), generates 177 valid strings, or in this case poison cans of food. Any food label that has at least one transition not included on the unsafe path map (i.e., contains at least one violation of the grammar) is sate to eat.

Model builders were shown a video tutorial on how to build an unsafe path map from examples of poison food labels. In the tutorial two valid strings were integrated to begin a transition map. The examples were selected to point out how to combine examples into a single map and to demonstrate the use of a loop that can create runs of the same letter in the strings. It is fairly easy to construct such a map with a reasonable level of accuracy. However, it is very difficult to get it exactly correct because one cannot always tell whether a letter represents returning to a previous transition or moving to a new node marked by the same letter (see Fig. 1). Thus, most of the maps created by our subjects were approximations to this actual transition map of the grammar. We gave our explicit groups a large set of strings (88 valid strings) to study, and ample time (three two-hour sessions) to learn something about the grammar. Implicit subjects were not aware of the existence of an unsafe path map. They simply tried to memorize the set of study strings. The implicit groups were similarly provided with equal time and cognitive artifacts (paper and pencil) to try to memorize the 88 study strings.

In each study session participants studied for 30 minutes (either memorizing strings or model building), then they took a 30 minute practice test. Half of the subjects in each learning mode (implicit or explicit) had a generate practice test and half had a string discrimination practice test. The generate practice test was just like the final test. Whenever a generated string had a reading of 7 (meaning it matched 70 percent of the letters in a valid string not previously generated) or higher, the closest valid string was shown on the computer screen and it was eliminated from the target set. However, only the 88 study strings were included in the target set for the practice test. That is, they were trying to generate only the study strings. The other 89 valid strings not in the study set were not recognized by the computer on the practice test. The string discrimination practice test consisted of presentations of a valid study string with a lure that contained one to four letters in places not allowed by the grammar (i.e., transitions not on the unsafe path map). The average number of violations in the lures was 2.4. Participants given this test selected which string they thought was poison (grammatical) on each trial and they received feedback on whether they were correct or not.

Upon reaching a criterion of 80% correct on their study list or after completing three study test sequences, learners attempted to generate all of the strings generated by the grammar (177 strings). It is this final string generation test that provided the measure of generative knowledge acquired by our subjects.

3.2. Results

The results for the participants who had the final discrimination test were: model builders with discrimination practice during training classified 83% of the strings correctly, model builders with generate training classified 85% correct. Memorizers with discrimination practice classified 74% correct and memorizers with generate practice classified 90% correct.

The results of most interest are those who had the final generation test. Initially we looked at the proportion of old (study list) strings generated separately from the proportion of new (never seen during training) strings generated. However, there were no differences in any of the conditions, so only the total number of strings generated will be presented. It should be noted however, that the lack of an old/new effect supports the notion that generative knowledge is being acquired in all our conditions. In other words, the memorizers were acquiring more than just memories for the strings presented during training.

Under these optimal conditions for reflective learning about the grammar, we expected our explicit subjects to quickly discover the rules of the grammar and outperform the memorizers (implicit groups) on the final generate test. Observations of the learners seemed to confirm our expectations. Participants in the explicit groups were more confident and seemed to enjoy the task much more than the memorizers (implicit groups). The model builders drew maps which became increasingly complex over sessions. When it came to the final test they confidently seemed to be using their maps to perform the test. The memorizers on the other hand seemed to get increasingly frustrated with the task. Eighty-eight strings seemed like too many to memorize and they all looked very similar. Memorizers spent a lot of time organizing the items for recall but did not seem to enjoy the task. We were quite surprised when we looked at the results of the final test.

On the final generate test the explicit learning groups with discriminate practice produced 45% of the valid strings (with the help of the computer's motherese), and explicit groups with generate practice produced 45% of the strings. Memorizers with discriminate training produced 38% of the strings and memorizers with generate practice produced 64% of the strings, significantly more than all the other conditions. In an ANOVA, the effect of type of practice test (discriminate versus generate) was significant, $F(1,87) = 6.85$, $p < 0.01$, as was the interaction of learning mode with practice task (discriminate versus generate training task), $F(1,87) = 5.53$, $p < 0.03$.

The above data reflect how good subjects were at producing or 'finding' the entire set of valid strings that can be generated by the grammar, It does not take into account the number of tries or number of strings generated by a subject. Recall that subjects were allowed to generate as many strings as they wished on the generate test, trying to find all the valid strings. The mean number of strings generated by the memorizers with discrimination training (117) was considerably lower than the other four conditions, who generated mean numbers of strings ranging from 159 to 171 strings. Therefore, we also looked at the data in terms of hit rate per attempt. This way of looking at the data produces a slightly different pattern: The model builders with discriminate practice had a hit rate of 45%. The model builders with generate training had a hit rate of 54%. The memorizers with discriminate training had a hit rate of 61% and the memorizers with generate training had a hit rate of 74%. The ANOVA on these data showed a significant effect of type of practice test, $F(1,87) = 5.06$, $p < 0.03$.

Thus no matter how you look at the data, the memorizers with generative practice are the best. The memorizers with discriminative practice come in second when looking at hits per try (a measure of quality of each generated string), but they come in last when looking at total number of poison cans found, because they produced fewer strings. Thus, in terms of ability to generate good strings (disregarding the number of strings produced), it appears that, contrary to our expectations, we replicated the frequent implicit learning finding: the implicit mode (memorizing strings) led to greater knowledge of the set of grammatical strings. That is, knowledge acquired implicitly about the grammar by memorizing strings was better able to support generating good strings than the explicit mode. This occurred in spite of our efforts to attach meaning to the strings, train subjects on how to build models of the grammar, allow them to use paper and pencil during training to develop a model of the grammar, and allow them to use their model on the final test. Also, this task requires actually generating the strings rather than the usual string discrimination test.

It should also be noted that, while building models of the grammar did not improve generativity, it did not greatly interfere with learning the grammar. The model building groups were only moderately inferior to the memorizing groups in hit rate per attempt, and they did better than the memorizers with discrimination training in terms of total number of valid strings produced in the final test.

Curiously, model building seemed to greatly enhance the confidence of the subjects in their ability to perform the task. They seemed to enjoy the task more and found it less frustrating. They also appeared to spend less time studying the strings. Thus we might speculate, in this paradigm, mental models served more of a motivational function than as a tool for classifying strings. We suspect that the same type of implicit knowledge used in the memorizing group was also at work in helping the model builders classify and generate strings because these subjects, like the memorizers, were exposed to the large set of valid exemplars during study. That is, when actually performing the task model builders may have relied more on their implicitly acquired knowledge than their mental model or cognitive artifact (their map of the grammar devised during training). We have seen this pattern of behavior before in other paradigms such as process control tasks, in which modification of learners' mental models of a task often have little effect on their performance (Mathews et al., 1996).

One more pattern in the generated strings of subjects in all the groups is quite interesting. If subjects were actually recalling stored memories of the valid strings to perform the generation task, we would expect that their sets of generated strings would begin with good ones — the ones they remembered well — and the quality of the strings they generated would deteriorate rapidly when they exhausted their list of remembered strings. Some slight decline would be expected across trials because as strings are generated they are removed from the target list. Thus, if participants accidentally generate strings they previously found later on the test, their quality score would go down because it would be rated as compared to the remaining closest valid string.

Actually, the pattern of quality of strings generated remained fairly constant across successive sets of generated strings. Fig. 2 shows string quality across quartiles of trials, using our ten point radioactivity (percent letters matching a valid string) measure. Since participants generated different numbers of strings on the generate test, we took each participants' data and divided the strings they generated into quartiles. The means plotted in Fig. 2 show the average quality of first through fourth quartile of strings generated by subjects in each group. As can be seen in Fig. 2, there is very little decline in quality of strings generated across trials. Interestingly, there

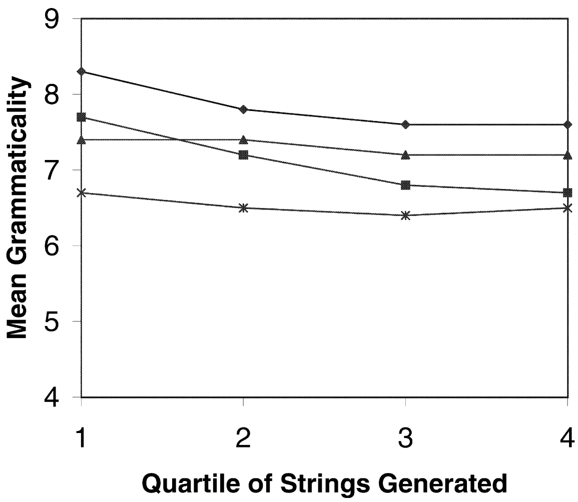

Fig. 2. Quality of strings as a function of quartiles of strings generated. (♦) Memorize/discriminate; (■) memory/generate; (▲) model/discriminate; (*) model/generate.

appears to be no difference in rate of decline as a function of model building versus memorizing. However, whereas the generate practice groups show virtually no decline, the discriminate practice groups show a slight decline over quartiles. However, the quality of strings generated remains high across all four quartiles. This pattern is consistent with the notion that subjects are not using rote memory of instances to generate strings. Instead they have learned about the grammar in terms of constraints and allowable transformations. Thus, then the string they generated follows pretty much the same constraints as the first several strings.

The results of this experiment support the conclusion that there is a similar level of generativity associated with implicit versus explicit instructions in this task. This result contrasts with findings in process control tasks that have found little evidence of generality in a task involving learning to control sugar production in a simulated sugar factory was above chance only in situations involving exact levels of workers used during training. Perhaps the difference in generalizability between the sugar task and the present experiment lies in the fact that in the grammar task subjects were exposed to a representative set of instances during training. The set of strings subjects were exposed to during training in these experiments may correspond to what Shaw, Wilson & Wellman (1986) have called a generator set. Thus, knowledge acquired about the nature of valid strings generalizes well to the complete set of strings generated by the grammar.

In the sugar production task, the set of situations experienced by a learner is under the learner's control. Each output level is dependent on the prior response of the learner. Thus, learners may adopt strategies that avoid sampling certain areas of the problem space (Buckner, Funke & Berry, 1995). However, one would still expect some transfer to new situations. Thus, this difference in generalizability of acquired knowledge may represent deeper differences between the knowledge used in the two tasks. Perhaps the knowledge acquired in process control tasks is limited to the episodic level (see Mathews & Roussel, 1998). However, the knowledge acquired in artificial grammar tasks seems to have very different qualities that enable relatively good generalization to new situations.

4. Experiment 2

In Experiment 1 our measure of generative knowledge of the grammar involved having subjects generate as many instances of the grammar (poison can labels) as possible. While this is a reasonable measure, it is different from what usually occurs in natural settings, such as the use of language. We rarely sit around trying to say everything we could possibly say in English! However, I think I have seen this behavior occasionally when students write papers for my classes. In the next experiment we developed a different measure of fluency that is more natural. We call it the cued generate task. In this task participants were given 3 randomly selected letters of a valid string (e.g.,- - - - S S - - - - V) and they were asked to type in the remaining letters that would make it a valid string. They were required to replace each blank in the cue with a single letter to create a valid string. The cues provide constraints on what subjects can generate. This is analogous to giving an appropriate reply in a specific context or to any situation where knowledge must be adapted to demands of constrain a subject to generating only one possible string. In most cases there are several valid strings that could be formed by completing the cues.

Two additional variables were examined in this experiment: number of exemplar strings seen during study and degree of freedom to organize the study strings. Number of exemplars seen during study is of interest for both practical and theoretical reasons. From a practical point of view, it is interesting to know if development of generative knowledge requires exposure to a large number of cases. From a theoretical point of view, this issue is related to whether generative knowledge depends on memorization of cases and analogical reasoning for applying the knowledge to new situations (called the exemplar view in the literature on implicit learning, e.g., Vokey & Brooks, 1992) versus learning to identify critical features or parts of strings and the constraints on how the parts can be combined to produce valid strings (e.g., Mathews, 1991; Perruchet & Pacteau, 1990; Perruchet, Vinter & Gallego, 1997; Servan-Schreiber & Anderson, 1990). If subjects are generating strings using analogies to specific cases, then the number of exemplars studied should be a very important variable. If they are learning about the structure of strings (e.g., how they are parsed and recombined) then only a minimal set of representative strings might be necessary. This variable was manipulated by providing subjects with either 22 study strings (short list) versus 88 strings (long list). The same grammar used in the previous experiment was used here. It can generate a total of 177 strings.

Freedom to arrange or organize study strings might be an important bridge between experiential (implicit) and reflective (explicit) knowledge. By sorting the strings into piles reflecting implicitly learned similarities (e.g., similar beginnings, endings or runs) one can bring into play reflective thought and make these regularities part of one's explicit model of the task. In the memory literature it has been demonstrated that organizational processes have a huge impact on memory (Tulving & Donaldson, 1972). Thus, freedom to organize strings may be an important way to get a synergistic interaction between implicit and explicit learning processes. This variable was maniulated by using either a bound set of cards with the study strings printed on them (fixed group) or loose cards that subjects were allowed to arrange as they wished during study.

4.1. Method

Seventy-nine undergraduate students participated in the experiment. All subjects in this experiment were given the implicit (memorize strings) version of the poison food game instructions. They were randomly assigned to four conditions reflecting all combinations of the two variables organizational freedom (free versus fixed study list) and study list length (22 strings or 88). This experiment consisted of a single session. Subjects had 20 minutes to study the strings and then they were given 20 minutes to work on the cued generate test.

On the cued generate test, each cue consisted of blanks for a string of a specific length with three letters filled in by the computer. The participant typed in letters in the blanks, attempting to find poison cans of food. If their string matched 70 percent of the letters in a valid string, the computer showed them the valid string and then a new cue was presented. Then, as in the earlier experiments, the valid string they generated was removed from the computer's target list so that the same string could not be generated again later. If they did not get close enough to the valid string, all of their wrong letters disappeared and they tried again. They had to keep generating strings using the same cue until they got close enough to a valid string. Only then would a new cue be presented. The program recorded the attempts necessary to successfully respond to each cue and the total number of valid strings generated in the 20 minute period.

4.2. Results

The mean number of valid strings generated in the four conditions were: 24.2 in the fixed, long study list group, 28.6 in the fixed short list group, 30.4 in the free long list group, and 31.2 in the free short list group. Thus, there was no evidence for any benefit of the long study list on total number of strings generated, There was a marginal benefit from being free to organize the study list, $F(1,75) = 3.747$, $p < 0.0567$. Although this effect was small, it should be remembered that this experiment involved only 20 minutes of study time. We think these data are encouraging that organizational processes might enhance development of generative knowledge over a longer period of study.

The fluency data (hits per attempt) present a different pattern. The means were 36.7% in the fixed long group, 36.2% in the fixed short group, 39.0% in the organize long group, and 57.1% in the organize short group. This interaction was highly significant, $F(1,75) = 8.989$, $p < 0.0037$. Clearly, the organize group with the short study list was much better than the other groups.

This experiment provides no evidence that exposure to a long list of exemplars is important in developing generative knowledge of the grammar. In fact exposure to a small set of representative cases with the opportunity to organize the cases led to greatly enhanced fluency in generating valid strings. Of course, this experiment is focused on the earliest part of development of generative knowledge. In the long run, over a much longer period of practice, we would expect exposure to a larger number of cases to be advantageous. However, these data support the notion that developing initial fluency in a new domain might be best approached by studying a small set of representative cases. The total lack of an effect of exposure to a very large number of cases might be of concern to those who propose that the primary learning mechanism involves storage of exemplars (e.g., Vokey & Brooks, 1992).

5. General discussion

Three changes were necessary to alter the artificial grammar paradigm to enable study of generative knowledge acquisition: (a) We made the task meaningful by equating strings with food labels that documented movements of the food around a ship and made the task more interesting by challenging the participants to find poison food cans. (b) We introduced a computerized form of 'motherese' that interpreted close but ungrammatical strings generated by participants and provided them with the correct string. (c) We exposed subjects to a representative set of strings before they attempted to generate strings. With these three elements in place it became possible to study generative knowledge of a complex grammar with as little as 20 minutes of study time prior to testing. We also developed the cued generate test that seems most similar to real world use of generative knowledge (Experiment 2). In this test subjects are constrained by the cues provided to generate certain types of strings, rather than attempting to generate all possible strings (Experiments 1).

Realizing that we are studying the early stages of generative knowledge acquisition, we found several factors that enhanced fluency development. Seeing a representative set of examples or cases prior to attempting to generate strings was crucial. In a prior experiment using the same poison food game and the computerized motherese, but not exposing subject to strings before beginning the generate task (Experiment 1 in Mathews & Cochran, 1998), subjects were not able to learn much about the grammar. They achieved a final hit rate of only about 5 percent and a much simpler grammar was used. In the Experiment 1 of the present study subjects in various groups generated 38 to 64 percent of the valid strings.

We also observed an advantage to letting subjects organize the set of study strings as compared to having the strings in a binder during study. Especially high performance was obtained with the short list (22 strings) and freedom to organize the strings during study. This finding suggests a type of synergistic interaction between experiential and reflective learning processes occurred. If learning were purely implicit, organizational processes would not be necessary and exposure to more cases (the long list) might be more advantageous. However, the fact that a shorter list of strings along with the opportunity to reflect about their relationships led to the best fluency suggests a positive interaction between the two types of learning.

We believe that processes like organizing the set of stimuli in one's own way during this early learning period help to develop bridges between the two types of knowledge being acquired about a domain. In other words fluency is maximized when experiential and reflective learning about the domain develop together in an interactive way. As many implicit learning studies have demonstrated, this is not always the case (see Reber, 1993 for a review). When dissociations occur it is usually the case that experiential knowledge is actually used to perform the task and mental models (reflective knowledge) are used to explain one's performance. Consequently, learners' descriptions of their own methods being used on the task may be very inaccurate. We suspect that organizational processes, like those used by memorizers in Experiment 2, may be useful tools for integrating the two types of knowledge and enhancing development of expertise. Perhaps this organizational process gives learners an opportunity to let their intuitive, implicit knowledge guide their classification and consciously observe the perceptual similarities and differences that are relevant to the domain. The potential use of organizational processes to integrate experiential and reflective knowledge should be further explored in future research.

Conscious efforts to discover the rules of the grammar has little effect on the nature or quality of generated strings even when subjects are trained in model building and are allowed to use cognitive artifacts (paper and pencil) during study. We had assumed that results from prior experiments showing that trying to figure out the rules of the grammar did not enhance performance was due to the fact that subjects could not find the rules. However, Experiment 1 demonstrated that even when people are trained in how to develop a map of the transitions in the grammar during study, and they are allowed to use their maps on the generation test, it did not enhance their performance. In terms of fluency or percent hits per attempt to generate a string, the memory subjects outperformed the explicit groups even under these extraordinary conditions. These findings highlight the fact that people acquire a lot of knowledge of new complex domains implicitly. Figuring out the rules or getting a conscious conceptual understanding of the domain follows rather than leads the process.

Mental models seemingly served primarily a motivational or emotional function. They seemed to act like a security blanket to prevent frustration with the immensity of the task. They gave subjects confidence, but had little impact on the quality or quantity of generated strings. This is not to claim that conscious reflective thought has no role in development of generative knowledge. We know from studies of expertise that deliberate practice, often involving application of reflective knowledge while performing the task, is essential for achieving top performance levels in any field (Ericsson, Krampe, & Tesch-Romer, 1993). As we said earlier, we believe that the development of generative knowledge involves a synergistic interaction between implicit and explicit learning processes. However, these data suggest that in the early stages of generative knowledge acquisition, seeing a representative set of cases is more important that drawing inferences about the domain. Generativity involves perception of relative invariance against a background of predictable changes. It requires representations that consist of reusable parts that are highly constrained in their ways of being recombined (Corballis, 1989). There is evidence of this type of generativity in perception (Biederman, 1986; Michaels & Carello, 1981), memory (Shaw, Wilson & Wellman, 1986; Wallace & Rubin, 1988), and implicitly acquired knowledge of grammars (Mathews & Cochran, 1998).

Two findings suggest that these results reflect true generativity in the sense of constrained adaptive responses rather than rote memory for instances. (a) The sets of strings generated by subjects tended to display a fairly flat progression in string quality across attempts to generate additional strings. If rote memory were being used we would have expected high quality of generated strings initially, followed by a drop in quality when subjects used up their list of well remembered instances. Instead, as in the case of memory for Ballads (Wallace & Rubin, 1988), the sequences of generated strings were relatively constant in quality, suggesting that they reflect knowledge of constraints (invariants) and allowable recombinations of reusable parts rather than memory of good cases followed by strategic guessing. (b) We also observed that having a study list of exemplars four times as large did not enhance learning. If sets of memorized instances was the key to implicit learning, we would expect a benefit from being exposed to the longer list.

We think this type of knowledge fits the description of Donald (1991) of mimetic level knowledge that marks the first point of departure in the evolutionary history of humans versus apes (see Mathews & Cochran, 1998). We conceptualize the mimetic controller as a presemantic perceptual representation system (e.g., Schacter, 1990, 1992a,b; Tulving & Schacter, 1990). It is similar to Mandler's concept (Mandler, 1992) of image-schemas. Its primary function is to build simplified conceptual representations that preserve temporal or spatial relationships. It produces nonlinguistic mental representations that provide us with information about object structure, location, and affordances (e.g., Schacter & Cooper, 1995). It has capabilities for limited mental inferences (e.g., object trajectory) and mental operations (e.g., delayed imitation of a rhythm). Thus the mimetic controller has some degree of internal structure or syntax that can be imposed on representations of stimuli such that different representations can be compared or combined. There is also evidence that the presemantic structural representation system (mimetic controller) is separate from symbolic or linguistic knowledge. For example, Warrington (1975) found that patients with dementia or severe visual object agnosia, who performed poorly on naming the functions of everyday objects, perform normally on tests of knowledge of object structure. Thus, mimetic representations seem to be the beginning of rule based representations that afford mental computation. Such internally constrained and recombinable representations appear to be what separates human from chimpanzee (Donald, 1991). Implicitly acquired knowledge of artificial grammars appears to be tapping into this basic level knowledge system that is the direct precursor of language.

References

Biederman, I. (1986). Recognition-by-components: A theory of human image understandinq. *Psychological Review 94*, 115–147.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science 16*, 153–184.

Buckner, A., Funke, J., & Berry, D. C. (Funke & Berry, 1995). Negative correlations between control performance and verbalizable knowledge: Indicators for implicit learning in process control tasks? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology 48A*, 166–187.

Cleeremans, A., & Jimenez, L. (1998). Implicit Sequence Learning: The truth is in the details. In: Stadler, M. A., & Frensch, P. A. (Eds.), Handbook of Implicit Learning, Sage, Thousand Oaks, CA.

Corballis, M. C. (1989). Laterality and human evolution. *Psychological Review 96*, 492–505.

Corballis, M. C. (1991). The lopsided ape, Oxford University Press, New York.

Donald, M. (1991). Origins of the modern mind: Three stages in the evolution of culture and cognition, Harvard University Press, Cambridge, MA.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (Krampe & Tesch-Romer, 1993). The role of deliberate practice in acquisition of expert performance. *Psychological Review 100-*, 363–406.

Degroot, A. D. (1966). Perception and memory versus thought. In: Kleinmuntz, B. (Ed.), Problem Solving, Wiley, New York.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review 99*, 587–604.

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., Buss, R. R., Chin, R., & Stanley, W. B. (Buss, Chin & Stanley, 1988). The role of explicit and implicit learning processes in concept discovery. *The Quarterly Journal of Experimental Psychology 40*, 135–165.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). The 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. (1998). Project Grammarama Revisited: Generativity of implicitly acquired knowledge. In: Stadler, M. A., & Frensch, P. A. (Eds.), Handbook of Implicit Learning, Sage, Thousand Oaks, CA.

Mathews, R. C., Lane, I. M., Roussel, L. G., Nagy, M. S., Haptonstahl, D. E., & Brock, D. B. (1996). Using conscious reflection, group processes, and Al to facilitate development of expertise. *Journal of Experimental and Theoretical Artificial Intelligence 8*, 259–276.

Mathews, R. C., & Roussel, L. G. (1998). Abstractness of implicit knowledge: A cognitive evolutionary perspective. In: Berry, D. (Ed.), How implicit is implicit learning, Oxford University Press, New York.

Michaels, C. F., & Carello, C. (1981). Direct Perception, Prentice Hall, Englewood Cliffs, NJ.

Miller, G. A. (1968). The psychology of communication: Seven essays, Penguin, Harmondsworth.

Nikolopoulos, C. (1997). Expert Systems, Marcel Dekker, New York.

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

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., & Gallego, J. (Vinter & Gallego, 1997). Implicit learning shapes new conscious percepts and representations. *Psychonomic Bulletin & Review 4*, 43–48.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and 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. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory 2*, 88–94.

Reber, A. S. (1993). Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious, Oxford University Press, New York.

Schacter, D. L. (1990). Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. *Annals of the New York Academy of Sciences 608*, 543–571.

Schacter, D. L. (1992a). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. *Journal of Cognitive Neuroscience 4*, 244–256.

Schacter, D. L. (1992b). Understanding implicit memory: A cognitive neuroscience approach. *American Psychologist 47*, 559–560.

Schacter, D. L., & Cooper, L. A. (1995). Bias in the priming of object decisions: Logic assumptions, and data. *Journal of Experimental Psychology: Learning, Memory and Cognition 21*, 768–776.

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

Shaw, R. E., Wilson, B. E., & Wellman, H. (Wilson & Wellman, 1986). Abstract conceptual knowledge: How we know what we know. In: McCabe, V., & Balzano, G. J. (Eds.), Event cognition: An ecological perspective, Lawrence Erlbaum, Hillsdale, NJ.

Simon, H. A., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology 5*, 29–46.

Stadler, M. A. (1995). The role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory and Cognition 3*, 674–685.

Stanley, W. B., Mathews, R. C., Buss, R. R., & Kotler-Cope, S. (Mathews, Buss & Kotler-Cope, 1989). Insight without awareness: On the interaction of verbalization, instruction and practice in a simulated process control task. *The Quarterly Journal of Experimental Psychology 41A*, 553–577.

Sun, R. (1995). Robust reasoning: integrating rule-based and similarity- based reasoning. *Artificial Intelligence 75*, 241–295.

Tulving, E., & Donaldson, W. (1972). Organization of Memory, Academic Press, New York.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science 247*, 301–306.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory and Cognition 18*, 328–344.

Wallace, W. T., & Rubin, D. C. (1988). ''The wreck of the old 97'': A real event remembered in song. In: Neisser, U., & Winograd, E. (Eds.), Remembering Reconsidered: Ecological and traditional approaches to the study of memory, Cambridge University Press, Cambridge, pp. 283–310.

Warrington, E. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology 27*, 635–657.