Psychologica Belgica
1997, 37-1/2, 69-88.

DOES THE ARTIFICIAL GRAMMAR LEARNING PARADIGM INVOLVE THE ACQUISITION OF COMPLEX INFORMATION?

Thierry MEULEMANS and Martial VAN DER LINDEN
University of Liège

The purpose of this article is to present a series of experiments originally aimed to study the respective role of fragment versus complex (rule-based) knowledge in artificial grammar learning, and to discuss some methodological constraints in artificial grammar learning tasks. What we are concerned with here is (a) the balance of different chunk-strength measures between grammatical and nongrammatical test items, (b) the importance of the number of items presented during the study phase, and (c) the importance of using a no-learning control group, with some insights on how these control subjects actually perform in the classification task. We also report an experiment conducted with the purpose to replicate the complex learning effect we observed in previous experiments, using a task in which grammatical and nongrammatical test items differed only according to their adherence to the rules. Finally, we present data obtained in young and elderly subjects, suggesting that implicit learning abilities are preserved in the elderly.

The artificial grammar learning paradigm was initially developed by Reber (1967) in order to demonstrate the ability of humans to abstract implicitly complex structural information from exemplars generated by a rule-governed system. This ability supposedly reflects very fundamental learning processes intervening in many different situations (e.g., social skill learning, language learning, emotional responsiveness, etc.). Many studies using different kinds of artificial grammar learning tasks have been carried out by Reber and his co-workers (see Reber, 1993, for a review) during the last three decades in order to investigate the implicit learning features.

In a typical artificial grammar learning task, subjects are first asked to memorize a series of letter strings generated by a finite-state grammar like the one illustrated in Figure 1. After this study phase, subjects are informed about the rule-governed nature of the study material, and are then presented with new letter strings, half of which are constructed with the same rules as the exemplars learned during the study phase, while the other half contain one or more violation(s) of these rules. The task is simply to classify these letter strings according to their adherence to the rules of the grammar. Typically, subjects do

Thierry Meulemans, Neuropsychology Unit, University of Liège, Belgium; Martial Van der Linden, Neuropsychology Unit, University of Liège, Belgium.

We thank Martin Redington for providing valuable comments on a previous version of this article.

Correspondence concerning this article should be addressed to Thierry Meulemans, Neuropsychology Unit, Faculté de Psychologie, Boulevard du Rectorat (B33), 4000 Liège, Belgium. Electronic mail may be sent to thierry.meulemans@ulg.ac.be.

not notice any rules during the study phase, and because they believe that they could not perform at an above-chance level, subjects are told that they can rely on their feeling, or intuition, to decide whether a letter string is grammatical or not. Of course, the interest of the task is that, despite their (apparent) guessing behaviour, subjects actually classify the test items at an above-chance level (with their performance generally reaching from 55 to 70% of correct responses).

These results have been confirmed in many studies since the late sixties; they also gave rise to many theoretical and methodological debates, especially since the eighties. One of the theoretical debates which is recurrent in artificial grammar learning studies concerns the opposition between the defenders of an abstractionist view of artificial grammar learning, on one hand, and a fragmentary/exemplarist view on the other. According to the abstractionist view, subjects base their well-formedness judgements in the classification phase on a knowledge acquired during the study phase, involving the abstraction of the rule-based structure of the grammar (see Reber, 1993). This abstracted rule-knowledge then permits the subjects to decide whether a new letter string corresponds to these rules or not. Note that, for many authors, this explanation is the most satisfactory to account for the transfer effects put forward in some artificial grammar studies: in these transfer situations, subjects are asked to classify new items constructed with another set of symbols than the ones used for generating the learning exemplars (e.g., a learning item could be XVVTM, and a test item could be DFFBG). About ten years ago, however, Reber's abstractionist view began to be challenged by researchers who argued that an abstraction mechanism was not necessary to account for subjects' performance in artificial grammar learning tasks (Dulany, Carlson, & Dewey, 1984; Perruchet, Gallego, & Pacteau, 1992; Vokey & Brooks, 1992). The main argument was that what subjects actually learn are the exemplars themselves, or fragments of exemplars, and that they base their classification judgments either on the similarity between test and study strings (the exemplarist view of Vokey & Brooks, 1992) or on a kind of familiarity effect founded on the relative frequency in the study items of string fragments present in the test items (e.g., Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990). According to the fragmentary view, subjects do not abstract anything from the learning items. What they learn are fragments (bigrams and/or trigrams), or chunks, of the letter strings. At the time of the test phase, the kind of information which could make them decide if an item is grammatical or not is whether this item is composed from chunks which were frequently seen during the learning phase or not. Actually, it seems that many results in artificial grammar learning studies can be accounted for by such a fragmentary view: most often, grammatical items contain more "familiar" (or less unfamiliar) chunks than nongrammatical items, and a knowledge about permissible bigrams and trigrams is sufficient to account for the subjects' performance (see Redington & Chater, 1996).

In that perspective, convincing evidence in favor of the fragmentary view has been recently reported by Knowlton and Squire (1994). These authors showed that, when manipulating independently the variables of global similarity between test and study items (global similarity being here defined by the presence, in a test string, of no more than one different letter from a study string) and of associative chunk strength (based on the frequency in the study strings of bigrams and trigrams appearing in the test strings), subjects based their classification judgments on the associative chunk strength (i.e., on the familiarity of string fragments) rather than on the global similarity. However, these results do not permit to decide whether the crucial factor in the test phase is the chunk strength of the test items or their adherence to the rules of the grammar. Indeed, in the Knowlton and Squire study, both variables were confounded. In a more recent study, Knowlton and Squire (1996) investigated whether subjects based their well-formedness judgments on the grammaticality or on chunk strength. The authors manipulated independently the two factors in their test items, and their results showed that both variables played a role in the grammaticality judgments.

Note that, although the debate between the abstractionist and the fragmentarist views seems to be the most prominent one in artificial grammar learning studies over the past ten years, other theoretical positions have recently been advanced. The most important one comes from the connectionist approach in cognitive psychology. Indeed, different connectionist models have been proposed to account for the subjects' performance in artificial grammar learning tasks (see Cleeremans, 1993; Dienes, 1992; Dienes, Altmann, & Gao, 1996). The main interest of these models is that, without involving any other learning processes than simple associative learning mechanisms, they seem to be able to learn the complex information present in stimuli generated by a finite-state grammar. With these models, it is no longer necessary to involve two different learning mechanisms (e.g., the first one being concerned with the learning of fragments, and the other with the abstraction of the grammatical structure): there is just a continuum from the learning of simple associations between elements (like bigrams) to the learning of the more complex conditional relationships between non-adjacent elements of the strings. Hence, connectionist models seem to be particularly well-suited to account for the learning mechanisms involved in implicit learning situations (and, more generally, in sequence learning).

Be that as it may, one of the major purposes in implicit learning studies is to investigate the learning of complex information. For that reason, it is important to ensure that the material used is able to show such a learning. With the artificial grammar learning paradigm, this involves the necessity to respect some methodological constraints, particularly concerning the construction of the test items. Indeed, if one wants to show a particular effect in an artificial grammar learning task, it is crucial to ensure that grammatical and nongrammatical test items differ only according to the variable being tested. For example, if the purpose of the experimenter is to show that his subjects learned some complex information during the study phase, he must be sure that his grammatical and nongrammatical test items do not differ according to more simpler information like their chunk strength characteristics.

The purpose of this article is to present a series of experiments originally aimed to study the respective role of fragment versus rule-based knowledge in artificial grammar learning, and to discuss some methodological constraints in artificial grammar learning tasks. What we are concerned with here is (a) the balance of different chunk-strength measures between grammatical and nongrammatical test items, (b) the importance of the number of items presented during the study phase, and (c) the importance of using a no-learning control group, with some insights on how these control subjects actually perform in the classification task. We also report an experiment conducted with the purpose to replicate the abstraction effect we observed in previous experiments, using a task in which grammatical and nongrammatical test items differed only according to their adherence to the rules (or to higher-level chunks than bigrams and trigrams). Finally, in order to show that implicit knowledge and explicit knowledge can be dissociated in an artificial grammar learning task, we present data obtained in young and elderly subjects, suggesting that implicit learning abilities are preserved in the elderly and showing that young adults perform better than elderly subjects in an explicit memory task.

Controlling the Chunk Strength

In two experiments conducted in the same perspective as the Knowlton and Squire (1994) study (Meulemans & Van der Linden, in press), we recently confirmed that subjects in an artificial grammar learning task could base their grammaticality judgments on their knowledge about fragments, and not on the adherence of the study strings to the rules of the grammar (the finite-state grammar used in these experiments is illustrated in Figure 1). Subjects were administered a task where the grammaticality and the chunk strength variables were balanced in the test items. Four groups of test items were constructed: grammatical and associated (GA), grammatical and nonassociated (GNA), nongrammatical and associated (NGA), and nongrammatical and nonassociated (NGNA). While the grammaticality variable reflected only the adherence of the items to the rules (a nongrammatical item violating the grammar at least in one position within the letter string), the associativity variable implied the same chunk strength measures as in the Knowlton and Squire (1994) study: [a] the global chunk strength (calculated, for each item, by averaging the different frequencies of its chunks - bigrams and trigrams - in the learning items), and [b] chunk strength for the anchor positions (calculated only for chunks located at the beginning and at the end of a letter string). In addition, we also controlled the "chunk novelty" variable (a "novel" chunk being a chunk which never appeared in the learning items). Hence, an associated item is a test item with a high global chunk strength, with a high chunk strength for the anchor positions, and without any novel chunk.

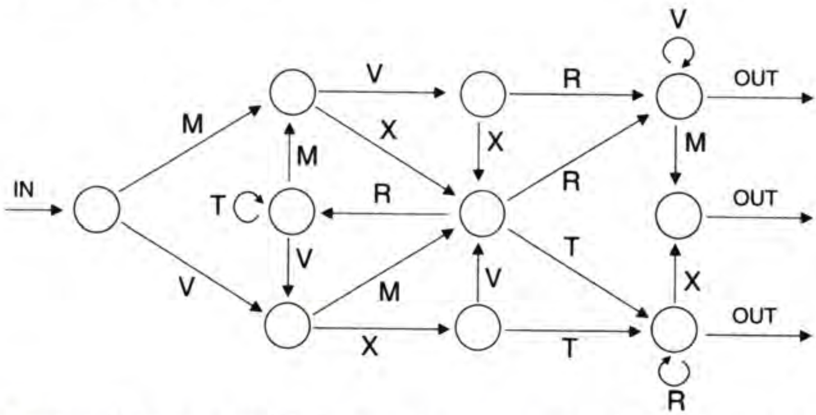

Figure 1. The finite-state Markovian rule system used in the Meulemans and Van der Linden (in press) experiments (taken from Vokey and Brooks, 1992). A finite-state grammar is composed of one or more entries and exits linked by multiple transition nodes. Letter strings are formed by starting at one entry, then by navigating from one transition to another, with each transition being able to generate a letter. A letter string terminates when an exit is reached.

Using this experimental design, it was possible to distinguish between the Reber abstractionist hypothesis and the fragmentary view of artificial grammar learning. If the adherence to the rules was the crucial variable, then subjects would endorse more often grammatical than nongrammatical items; on the contrary, if subjects based their judgments on some kind of knowledge about fragments of the items, then they would accept as grammatical the associated items more often than the nonassociated ones.

Results showed a main effect of the associativity variable, and no effect of grammaticality. Hence, and contrary to results obtained in many studies on artificial grammar learning, our subjects were not able to distinguish between grammatical and nongrammatical items, but based their judgments on the associativity variable. This result is important, because it questions previous evidence which had suggested that subjects could correctly classify letter strings on the basis of their adherence to the grammatical rules; it could be that, at least in some cases, this result was due to the lack of control of the bigrams and trigrams familiarity in the test strings. Actually, if this variable is not controlled in the test items (which can be done by using the procedure of Knowlton and Squire, 1994, that is, by calculating the associative chunk strength), then grammatical test items tend "naturally" to contain more associated chunks than nongrammatical items do.

However, our study also suggested that controlling the associative chunk strength, as Knowlton and Squire did, may not be sufficient to completely eliminate the possibility for the subjects to rest their judgments on some chunk-strength based knowledge. Indeed, we showed that the presence, in a particular test item, of one or more chunks which were not present in the learning items could lead to a non-familiarity feeling for that item, whatever its global chunk strength. Likewise, it has been shown that the chunks located at the beginning and at the end (anchor positions) of the strings could be of particular importance in the grammaticality judgments (Perruchet & Pacteau, 1990; Reber & Lewis, 1977).

To summarize, it appeared that global chunk strength, chunk strength for the anchor positions, and the chunk-familiarity variable should always be controlled in artificial grammar learning studies. Actually, this should be done separately for the bigram and trigram frequencies, and not only by averaging their frequencies (see Redington & Chater, 1996, for discussion).

No Possible Abstraction?

In the same study (Meulemans & Van der Linden, in press), we also addressed another question, namely, whether the absence of a grammaticality effect stated above could be due to another factor, which is the number, or the proportion, of grammatical items presented during the learning phase. The question was: how could a rule-abstraction effect be observed in a situation where the subjects are confronted with only a few exemplars of the grammar? Actually, in the preceding experiments in which the chunk strength measures were balanced across grammatical and nongrammatical items, subjects learned respectively only 16 and 32 grammatical items in the study phase, while the grammatical structure could produce 150 different strings comprising from 3 to 7 letters. Hence, it could be that this number of exemplars was too small to embody a sufficient part of the grammatical structure, and that subjects could only base their judgments on superficial aspects of the learning strings.

Controlling the chunk strength measures in the same manner as described above, subjects were presented with 125 items at the learning phase, that is to say, a much larger proportion of the items the grammar could generate. Results showed the inverse pattern of the previous experiments: grammatical items were endorsed more often than nongrammatical items, whatever their associative chunk strength. This could not be attributable to a better explicit knowledge gained during the task: indeed, there was no correlation between the subjects' performance in the classification task and their performance in a generation task in which they were asked to generate ten letter-strings they thought were grammatical. Actually, a comparison between the two conditions (few exemplars vs. many exemplars during the learning phase) revealed that levels of explicit knowledge were not higher in the second than in the first condition. This result could not be attributable to the lack of sensitivity of the generation task, because subjects who did not benefit from a learning phase (control condition) performed worse than experimental subjects in the generation task.

The Importance of a No-Learning Control Group

Some authors have argued for the necessity to include a control condition in every study on artificial grammar learning (e.g., Redington & Chater, 1996). In that control condition, subjects would have irrelevant training, or they would only be submitted to the classification phase, without any preliminary confrontation with exemplars of the grammar. If we want to ensure that the ability to classify grammatical and nongrammatical strings is actually based on some kind of knowledge (either abstract or exemplar-based) acquired during the study phase, we should compare this performance with that of a no-learning control group. Moreover, it has been hypothesized that subjects in artificial grammar learning tasks would actually learn to identify grammatical and nongrammatical strings during the classification phase itself, and not during the study phase. The systematic inclusion of a no-learning control condition (or a group with no relevant training) in artificial grammar learning tasks would be the only way to make sure that performance is actually based on the study phase, and, hence, is crucial for any demonstration of implicit learning in artificial grammar learning.

In two experiments (Meulemans & Van der Linden, in press), we included such control conditions, and demonstrated that performance of the experimental subjects in both conditions (few exemplars vs. many exemplars during the study phase) was determined by knowledge acquired during the study phase. However, we also showed significant effects in the control groups. In one experiment, and although this effect was not as strong as in the experimental group, control subjects were also sensitive to the associativity variable. The other experiment showed a significant grammaticality effect in the control group, but in the inverse direction: control subjects endorsed more often nongrammatical than grammatical strings.

At first glance, these effects appeared to be quite surprising. How could control subjects decide, otherwise than by chance, whether nonsense letter-strings were correct or not according to a set of rules with which they had never been confronted? In fact, in studies in which control groups were used, contradictory results were obtained, sometimes not significantly differing from chance (Altmann, Dienes, & Goode, 1995; Perruchet & Pacteau, 1990), and sometimes showing above-chance performance (Dulany, Carlson, & Dewey, 1984). Dulany (cited in Redington & Chater, 1996) suggested that this effect could be due to the presence of familiar letter strings (such as MTV, for example) in the grammatical test items set. On the other hand, according to Redington and Chater, one cannot reject the possibility that control subjects could learn to differentiate grammatical from nongrammatical strings during the test phase. Moreover, these authors suggest that the contrasting evidence in the literature could also be accounted for in terms of motivational or other methodological differences (e.g., the particular instructions given to control subjects at the beginning of the test phase) between experiments.

However, in our study, the same methodology was used in both experiments, and no motivational differences between control groups can be suspected (the only difference between the two experiments being the length of the learning phase). We therefore re-analyzed our test items according to their "internal" chunk strength, that is, chunk strength calculated not with regard to the learning strings, but to the frequency of their constituent bigrams and trigrams in the test strings themselves. This internal chunk strength effect would develop progressively during the test phase, as the subjects are confronted with grammatical and nongrammatical items. In fact, because the internal chunk strength properties of the test items were never controlled in artificial grammar learning tasks, their means are likely to be differentially distributed between grammatical and nongrammatical items across studies. In other words, internal chunk strength may be either higher or lower for grammatical strings than for nongrammatical strings, or equal for both types of items.

In the experiment where control subjects were sensitive to the associativity variable, we showed that internal chunk strength of associated items was significantly higher than that of nonassociated items. On the other hand, the inverse grammatical effect observed in the other experiment could be explained by the higher internal chunk strength of the nongrammatical strings. Although this should be verified for the test items in every study in which a control group is used, it is likely that at least some of the contradictory data reported (Redington & Chater, 1996) could be explained by this internal chunk strength effect as well. However, in some cases, other factors must also be involved, because some studies in which the same material was used showed different patterns of results in control subjects (Altmann, Dienes, & Goode, 1995, Experiment 1 & 2; Dulany et al., 1984; Perruchet & Pacteau, 1990; Redington & Chater, 1994). If the only variable determining grammaticality judgments of control subjects was the internal chunk strength of the test items, there should have been no difference between these studies. It is therefore necessary to involve other explanations, which remain to be found (some proposals of Redington & Chater could be potential candidates).

The internal chunk strength effect could also have played a role in the recent study conducted by Knowlton and Squire (1996, Experiment 1), who compared performances in amnesic patients and normal subjects in an artificial grammar learning task. Knowlton and Squire showed that both amnesic patients and normal subjects were able to classify grammatical and nongrammatical items, and that there was also an effect of the chunk strength variable, indicating that letter strings with a high chunk strength were endorsed more often than letter strings with a low chunk strength. And finally, there was an interaction between the grammaticality and chunk strength variables, showing that the chunk strength effect was mostly important for nongrammatical strings. However, no control group was included in the Knowlton and Squire experiment. Reanalyzing their test items according to their internal chunk strength, we showed a close parallelism between the pattern of results in the Knowlton and Squire study and the distribution of the internal chunk strength means across types of test items (see Figure 2). Hence, it could be that their results in amnesic patients were biased by this lack of control of the internal chunk strength of the test

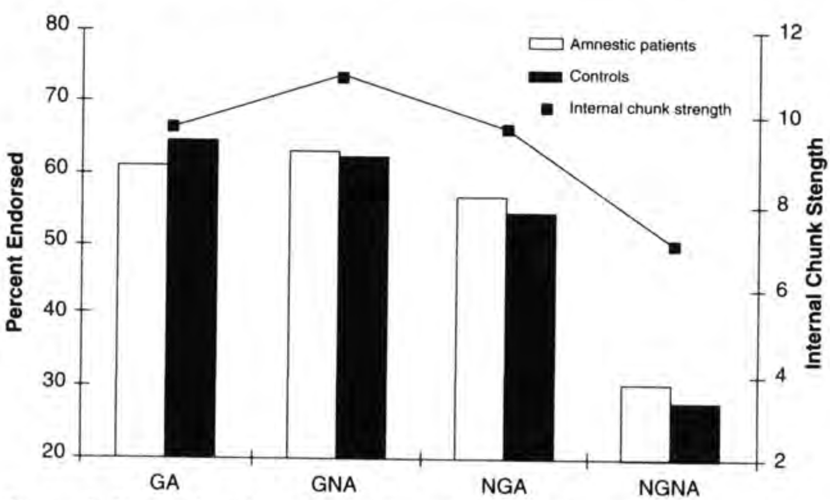

Figure 2. Comparison between internal chunk strength and endorsement rate for normal participants and amnesic patients in the Knowlton and Squire (1996, Experiment 1) study. Endorsement rates are taken from Knowlton and Squire's Table 3 (page 173). The open bars show endorsement rates for amnesic patients and normal subjects (left vertical axis), and the lines show the internal chunk strength values for each item type (right vertical axis).

items. In that perspective, it could be hypothesized that severe amnesic patients are in fact in a situation similar to that of no-learning control subjects. At present, one cannot reject the idea that these patients did not learn anything from the study phase, and that they could actually only base their classification judgments on the internal chunk strength of the test items. Note that this assumption is also valid for transfer situations, in which subjects have to classify test items constructed with another set of symbols than those used for the learning items. Without the inclusion of a no-learning control group, transfer effects remain questionable as well.

A Replication

Respecting the different methodological points mentioned above, we devised a new artificial grammar learning task in order (1) to replicate, with another finite-state grammar, results obtained in our recent study showing a grammaticality effect and (2) to explore the implicit learning abilities in elderly subjects and brain-damaged patients. The choice of the finite-state grammar was determined by the number of strings from 4 to 7 letters it could generate, because we wanted both to present the majority of the grammatical strings during the learning phase and to construct a task which could be used in the clinical domain (i.e., not too long).

Twenty university undergraduates (10 males and 10 females) aged between 19 and 29 years (mean age = 22.8) volunteered for the experiment and were tested individually (experimental condition). Another group of 20 young subjects (10 males and 10 females), aged between 20 and 30 years (mean age = 24.2) participated in the no-learning control condition.

The finite-state grammar (see Figure 3) was adapted from Mathews, Buss, Stanley, Blanchard-Fields, Cho, and Druhan (1989). The purpose was to obtain a grammatical structure which could produce about 60 letter strings from 4 to 7 letters.

The letter strings were presented, one at a time, on 9 cm x 14 cm index cards, in a predetermined random order. Experimental subjects were administered the learning phase, followed by the classification phase and by the explicit test. Control subjects were only administered the classification phase.

Learning phase and classification phase. Firstly, all the possible letter strings containing from 4 to 7 letters were generated. Out of these 63 grammatical strings, two groups of grammatical strings were constructed: a first group of 51 strings for the learning phase, and a second group of 12 strings for the test phase (see Appendix). In our previous experiments (Meulemans & Van der Linden, in press), 16 grammatical strings were selected for the test phase, and 125 items out of the 150 grammatical ones were presented at the learning phase, which

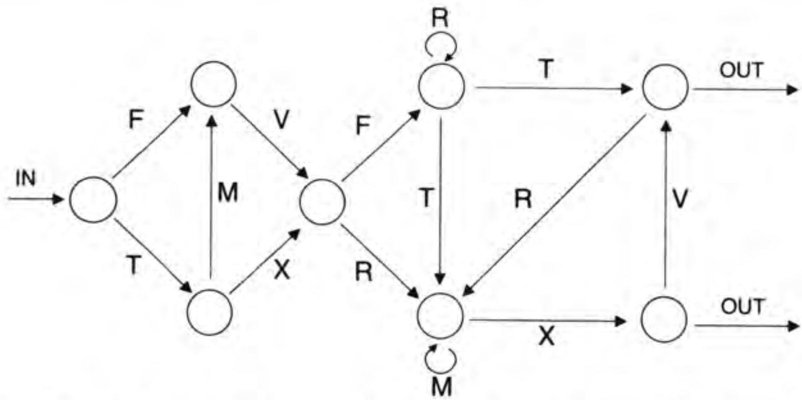

Figure 3. The finite-state grammar used in the present experiment (adapted from Mathews et al., 1989).

represents a ratio of 83.3% of grammatical strings. In the present experiment, in order to maintain a sufficient ratio of learning strings, the number of test items was reduced to 12, and the 51 learning strings represent 80.95% of the entire grammatical strings series.

Another series of 12 nongrammatical strings was constructed for the test phase. These nongrammatical items were generated by violating the rule system at one or two positions within the letter string. These rule violations could occur in any position within the letter string, except in the first and last position. The associative chunk strength and chunk-novelty were controlled in the grammatical and nongrammatical test items. The associative chunk strength of a test item was determined on the basis of the frequency with which chunks (bigrams and trigrams) making up the test strings appear in the learning strings (see Knowlton & Squire, 1994; Meulemans & Van der Linden, in press). Two indicators of associative chunk strength were used: the global associative chunk strength (for chunks located at any position within the letter string), and the chunk strength for the anchor positions (for chunks located at the beginning and at the end of the letter string). The chunk-novelty variable concerned the presence in the test items of chunks never encountered in the learning items. Again, two indicators of chunk-novelty were used: the global chunk-novelty, indicating the percentage of novel chunks in the test item, and the chunk-novelty for the anchor position, indicating the number of novel chunks located at the beginning and at the end of the string (see Meulemans & Van der Linden, in press, Experiment 2a, for details concerning this procedure).

These four indicators (associative chunk strength and chunk-novelty, both for chunks located at any position and for chunks located in the anchor positions) were balanced through the grammatical and nongrammatical test items (see Table 1 of the Appendix for details). The mean associative chunk strength was the same for grammatical and nongrammatical items (12.41 vs. 12.93), $t(22) = -.77, p > .40$, as was the mean associative strength for the anchor positions in grammatical and nongrammatical items (10.88 vs. 10.79), $t(22) = .15, p > .80$. Likewise, the global chunk-novelty and the chunk-novelty for the anchor position were the same in grammatical and nongrammatical items (.50 vs. .67 for the global chunk-novelty, $t(22) = -.80, p > .40$; none of the test items contained a novel chunk located at the anchor position).

In the learning phase, the task was presented as being an immediate-memory test. Each index card was presented for 3 sec., and then the subject was asked to reproduce it orally. If the response was correct, the next string was presented in the same way. If the subject was unable to repeat the letter string correctly, the same string was presented a second time. After a maximum of three

Table 1
Chunk Strength Values of Test Items

Test items	Grammatical status	Global chunk strength	Anchor position chunk strength	Chunk novelty
TXFTRMX	G	11,27	11,75	1
FVFT	G	16,60	9,00	0
FVFRTMX	G	12,45	12,25	1
FVFRTRX	G	10,19	11,00	1
FVRMXV	G	11,44	11,75	0
FVFTRMX	G	11,45	12,25	1
TXFRTRX	G	10,54	10,50	1
TMVFTX	G	15,00	8,00	0
TXFRRT	G	11,56	10,50	0
TXFRTMX	G	12,09	11,75	1
TXFRTX	G	14,11	10,25	0
TXRMXV	G	11,44	11,50	0
TXFRMXV	NG	11,45	12,75	1
TXFVRX	NG	13,00	10,75	1
TXRTXV	NG	12,78	11,50	1
TMVRTXV	NG	12,09	11,00	1
TXFTXFT	NG	15,18	8,50	0
TXFTXRX	NG	13,45	10,00	0
FVFTXRX	NG	13,64	10,50	0
FVFVRX	NG	12,89	11,25	1
FVFTXFT	NG	15,36	9,00	0
TXRRT	NG	10,14	9,25	1
FVFRXV	NG	12,67	13,25	1
FVRTXV	NG	12,44	11,75	1

presentations, the examiner went on to the next item. The 51 letter strings were presented in a predetermined random order for half of the subjects, and in the inverse order for the other half of the subjects. In the learning phase, subjects were able to reproduce correctly 82.1% of the letter strings in the first trial; 94.2% of the items were correctly reproduced after two trials, and 96.5% after three trials.

Five minutes after the end of the learning phase, the subjects were instructed that the letter strings they had just memorized a few minutes ago were constructed according to a complex system of rules. They were asked to classify new items as grammatical or not. The test items were presented in the same way as in the learning phase, except that there was no presentation time limit. There was no feedback given as to the correctness of the judgments. The 24 letter strings were presented in a predetermined random order for half of the subjects, and in the inverse order for the other half of the subjects. The procedure was repeated, with the same letter strings being presented a second time in the same order.

A word-generation task was administered at the end of the classification task in order to assess the explicit knowledge gained by the subjects. They were given a sheet of paper where the six letters contained in the grammar were printed; underneath these six letters, they were asked to write 10 strings they considered to be grammatical.

Results. The percentage of grammatical strings classified as grammatical by the experimental subjects was 65.0% (SE_M = 2.4%), and 52.3% (SE_M = 3.72%) for the nongrammatical strings. Subjects in the no-learning control condition classified as grammatical 47.9% (SE_M = 3.8%) of the grammatical items, and 53.5% (SE_M = 3.4%) of the nongrammatical items (Figure 4). An ANOVA with Group (experimental vs. control) as between-subject variable and Grammaticality (grammatical vs. nongrammatical) as repeated-measure factor showed an effect of Group, $F(1,38) = 4.40$, $MS_e = 284.07$, $p < .05$, indicating that experimental subjects accepted more items as grammatical than control subjects, and a significant interaction between Group and Grammaticality, $F(1,38) = 9.75$, $MS_e = 172.72$, $p < .005$, indicating that (Newman Keuls post hoc test) experimental subjects endorsed more often grammatical items than nongrammatical items ($p < .05$), and that control subjects did not ($p > .30$).

Contrary to subjects in the experimental condition, the endorsement rate in the control condition was significantly correlated (r = .74; for experimental subjects : $r = -.31$) with the internal chunk strength of the test items (i.e., chunk strength calculated not on the learning items, but on the test items themselves). This result, confirming previous data obtained in our laboratory (Meulemans & Van der Linden, in press), suggests that a no-learning condition cannot be considered as a kind of 'neutral' condition, in which subjects would respond at random.

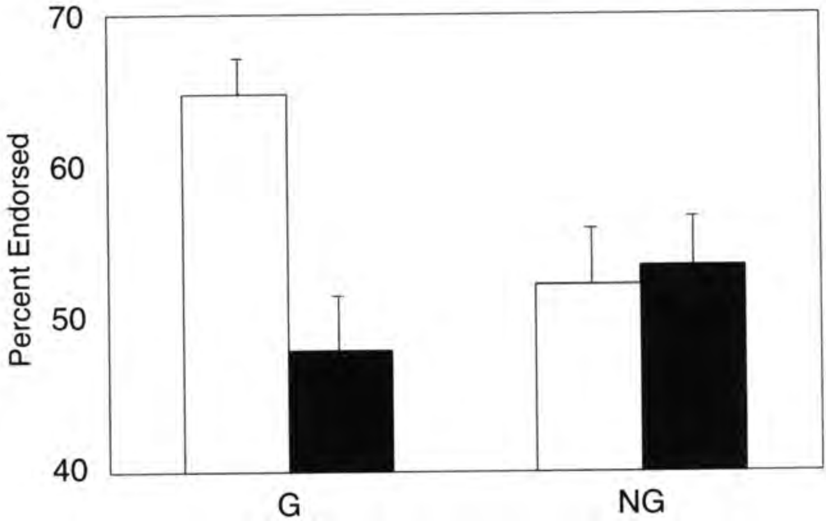

Figure 4. Percentage of items endorsed as grammatical. The open bars show performance for experimental participants, and the shaded bars show performance for controls. Error bars show standard errors of the mean.

A correlation analysis was computed between the percentage of correct responses for each experimental subject in the classification task and the number of grammatical strings written by each subject in the generation task. The correlation was not significant (r = -.10). We also analyzed the items according to their associative and anchor position chunk strength. There was no significant correlation between the associative chunk strength of the strings generated by each subject and the percentages of correct responses in the classification task (r = -.07 for global associative chunk strength, r = .23 for chunk strength at the anchor positions).

To summarize, the purpose of the present experiment was to replicate, using a different finite-state grammar, results obtained in a preceding study which suggested that it is possible to show a grammaticality effect in a well-controlled artificial grammar learning task by presenting a sufficient proportion of exemplars of the grammatical structure during the learning phase. Results showed that subjects were able to discriminate between letter strings which differed only according to their grammatical status, and not according to other more superficial features like, for example, their bigram and trigram frequencies.

In this study, and contrary to the our previous experiments, both grammatical and nongrammatical items did not differ with regard to their internal chunk strength (calculated on the basis of the frequencies, in the test strings series, of their bigrams and trigrams). For this reason, and as we expected, the no-learning control subjects in the present experiment could not discriminate between grammatical and nongrammatical items on that basis, and their performance was not better than chance. However, correlation analysis confirmed the importance of the internal chunk strength variable for the control subjects: there was a significant correlation between the internal chunk strength of each test item and its endorsement rate by control subjects, while this was not true for experimental subjects. These results confirm (a) that a no-learning control condition cannot be considered as being a neutral situation, and (b) that the performance of experimental subjects was determined by the learning phase and not by some features of the test items themselves (see Redington & Chater, 1996).

An Application

The artificial grammar learning task described above appears to be a well-controlled task, potentially usable in order to explore rule-abstraction abilities in different groups of subjects presenting memory deficits. So, we decided to conduct a first study with the purpose to assess implicit learning abilities in a group of elderly subjects.

It is well documented that elderly adults perform poorly on explicit tests of memory, such as recall or recognition, which require a deliberate recollection of specific episodes. On the other hand, when young and elderly adults are compared on implicit memory tasks (e.g., perceptual priming tasks), age differences are generally small and unreliable (La Voie & Light, 1994; Rybash, 1996). In the same perspective, it has been argued that implicit learning should be preserved with age (e.g., Reber, 1993). However, studies showing the absence of any age effect on implicit learning abilities are relatively sparse, and none was conducted with the artificial grammar learning paradigm (see Cherry & Stadler, 1995; Curran, in press; Howard & Howard, 1992). Therefore, the statement of the absence of any age effect on implicit learning abilities remains largely hypothetical.

We administered the task described above to a group of 20 young (aged between 21 and 26 years; mean age = 23.3) and 20 elderly subjects (aged between 60 and 69 years; mean age = 64.6); the two groups were matched on the basis of their educational levels (16 years of education on average for both groups). Results showed that elderly subjects performed as well as young subjects in the classification task: the percentage of grammatical strings classified as grammatical by the young subjects was 67.1% (SE_M = 3.1%), and 58.4% (SE_M = 2.8%) for the nongrammatical strings; elderly subjects classified 61.9% (SE_M = 2.4%) of the grammatical strings and 52.3% (SE_M = 2.3%) of the nongrammatical strings as grammatical.. Results also confirmed the dissociation between implicit and explicit measures of performance. Contrary to the classification task, young subjects performed better than elderly subjects in the generation task when considering the mean chunk strength of the items they generated (7.68 vs. 6.50, $t(38) = 2.56$, $p < .05$). Moreover, there was no correlation between performance in the generation task and performance in the classification task. We think that these results are important: the fact that young subjects performed significantly better than elderly subjects in the generation task supports the idea that the generation task can actually be considered as a sensitive test of explicit knowledge. Indeed, the age effect on explicit memory tasks is well-known: young subjects typically perform better than elderly subjects on explicit long-term memory measures like free recall and cued recall (see Van der Linden, 1994). Therefore, a measure supposed to tap the explicit knowledge of the subjects in an artificial grammar learning task has also to be also sensitive to age differences. Our result confirms that this was the case with the generation task. This result also suggests that the two tasks (classification task and generation task) actually tap independent knowledge bases.

Conclusion

In this article, we have tried to focus on the importance of some methodological constraints in artificial grammar learning tasks. We have presented a series of experiments originally aimed to clarify how processes based on superficial features of the learning strings versus abstraction mechanisms can account for performances in the grammaticality judgments. More specifically, our results suggested that, under certain conditions (e.g., when the proportion of items presented during the study phase is weak), grammaticality judgments can be influenced by some kind of chunk familiarity, but also that this effect can be suppressed in favor of a rule-abstraction mechanism when more items are presented during the learning phase. It could be that the crucial point is, in fact, to ensure that the learning string series is representative enough of the grammatical structure. It can be expected that this can be more easily achieved when many different items are presented during the learning phase than when the number of learning items is weak (see also Shanks, Johnstone, & Staggs, 1997).

However, and although our results can be interpreted in terms of the fragment vs. rule learning dichotomy, we suggest that our data can be better accounted for by another theoretical interpretation, which comes from works realized from a connectionist point of view (e.g., Cleeremans & McClelland, 1991). Indeed, Cleeremans et McClelland have shown how a connectionist network could simulate the performance of normal subjects in a probabilistic sequence learning task (which is, in many aspects, comparable to the artificial grammar learning task). More particularly, they showed how a Simple Recurrent Network (S.R.N.) can be sensitive to the elementary characteristics of the stimuli (frequency of isolated elements, simple associations) as well as to more complex associations between elements. This result is important, because it shows how a single mechanism based on associative learning processes can account for the apparent dissociation put forward in our experiments.

However, we still believe that the methodological questions developed in the present article are relevant. It must be recalled that one of the main purposes of implicit learning studies is to show the ability to learn complex information. If the subjects' performance in a particular artificial grammar learning task can only be explained either by a knowledge about chunks or by simple associations between adjacent elements of the material, one should not talk about "complex learning". Therefore, it is necessary to control rigorously the test items in order to ensure that the ability of the subjects to classify grammatical and nongrammatical strings is only due to the learning of complex information in the study material. More specifically, our results suggest that, in order to put forward the subjects' ability to classify letter strings on the basis of their knowledge of complex conditional associations between elements, different chunk strength measures must be balanced (for bigrams and trigrams) across grammatical and nongrammatical test items: global chunk strength, chunk strength for the anchor positions, and the presence of novel chunks within the letter strings. Without taking these methodological precautions, it would be virtually impossible to argue against the Redington and Chater (1996) contention that evidence in favor of abstract knowledge can almost always be accounted for by fragment knowledge as well.

Our data also give some insights on what control subjects not previously confronted with the grammar do when they are asked to classify grammatical and nongrammatical items. It appeared that the situation of these control subjects is not a situation where responses are made at chance, and also that the differential results between experimental and control subjects cannot be accounted for by some kind of motivational differences between the two groups. In fact, controls also base their judgments on some kind of fragment knowledge, but this knowledge concerns, for each test item, the frequency of its chunks in the whole test item series.

And finally, we have shown, using a task in which grammatical and nongrammatical items did not differ according to their chunk familiarity features, that elderly subjects could classify test items significantly above chance, and that their performance reached the same level as that of young adults subjects. Thus, it appeared that implicit complex learning abilities in artificial grammar learning are preserved in the elderly.

Considering the methodological points raised above, many results in artificial grammar learning studies can probably be questioned. In particular, recent data collected by Knowlton and Squire (1996) in amnesic patients are flawed by a lack of methodological controls. Actually, no published evidence suggesting preserved implicit learning abilities in amnesic patients using the artificial grammar learning paradigm are, to date, really convincing. We have conducted (Meulemans, 1997), in our laboratory, a study aimed to investigate implicit learning abilities in amnesic patients with the task described in this article. We obtained the same pattern of results as in our study with elderly subjects: there was no significant difference between amnesic and matched normal subjects in the classification task, but the two groups differed in the explicit generation task.

We also investigated implicit learning abilities in patients suffering from lesions of the striatum. It has often been shown that these patients present abnormal performances in procedural skill learning. One of the interests of studying implicit learning in such patients is to try to obtain some insights with regard to the distinction between implicit and procedural learning. The fact that both kinds of learning are underlain by the same brain structures (i.e., the basal ganglia) would be an argument (even if not critical) in favor of the similarity between the two concepts. On the contrary, if patients suffering from lesions to the striatal structures perform normally in implicit learning tasks, then one should consider the possibility that procedural learning tasks and implicit learning tasks are underlain by separate learning mechanisms. In that perspective, we have conducted a study with Parkinson patients (Meulemans, 1997), whose implicit learning abilities were to date and to our knowledge never investigated with the artificial grammar learning paradigm. Our results showed normal implicit learning with the artificial grammar learning task in a group of 17 non-demented Parkinson patients (note however that there was a numerical difference between the results of Parkinson patients and normal subjects, and the absence of significant difference could be due to a lack of statistical power).

References

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 899-912.

Cherry, K. E., & Stadler, M. A. (1995). Implicit learning of a nonverbal sequence in younger and older adults. *Psychology and Aging, 10,* 379-394.

Curran, T. (in press). Effects of aging on implicit sequence learning: Accounting for sequence structure and explicit knowledge. *Psychological Research.*

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

Howard, D. V., & Howard, J. H. (1992). Adult age differences in the rate of learning serial patterns: Evidence from direct and indirect tests. *Psychology and Aging, 7*, 232-241.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 79-91.

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

Knowlton, B.J., Squire, L.R., & Butters, N. (1994). Intact artificial grammar learning in Huntington's disease. *Society for Neuroscience Abstracts, 20*, 1075.

La Voie, D., & Light, L. L. (1994). Adult age differences in repetition priming: A meta-analysis. *Psychology and Aging, 9*, 539-553.

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

Meulemans, T. (1997). *L'apprentissage implicite: Une approche cognitive, neuropsychologique et développementale*. Unpublished doctoral thesis, University of Liège, Belgium.

Meulemans, T., & Van der Linden, M. (in press). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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., Gallego, J., & Pacteau, C. (1992). A reinterpretation of some earlier evidence for abstractiveness of implicitly acquired knowledge. *Quarterly Journal of Experimental Psychology, 44A*, 193-210.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 77*, 317-327.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Reber, A. S., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition, 5*, 333-361.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General, 125*, 123-138.

Rybash, J. M. (1996). Implicit memory and aging: A cognitive neuropsychological perspective. *Developmental Neuropsychology, 12*, 127-179.

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

Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology, 50A*, 216-252.

Van der Linden, M. (1994). Mémoire à long terme et vieillissement. In M. Van der Linden & M. Hupet (Eds.), *Le vieillissement cognitif* (pp. 87-140). Paris: Presses

Universitaires de France.
Vokey, J. R., & Brooks, L. R. (1992). The salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 328-344.

Appendix

Learning items

TXRXV	TXRXVRX	FVFTRXV
TXFRRRT	FVFTRX	FVRMMXV
FVFTMXV	TMVRX	TXFT
TXFTRXV	TMVRMXV	FVRX
TMVRMX	FVRMMMX	TMVFRRT
TMVFRTX	TMVFTRX	TXFTXV
TXFRRTX	TXFRT	TMVFTXV
FVFRRRT	TXFTMXV	FVFRRT
TXRX	FVFTMX	FVFRTX
TXFTMX	FVRMX	FVFTX
FVFTXV	TXRMX	TXFTX
TMVFT	FVFRRTX	TXFTMMX
FVFRT	FVRMMX	TMVFTMX
TXRMMX	FVFRTXV	TXFRTXV
FVFTMMX	TXRMMXV	TMVFRT
TMVRMMX	TXFTRX	FVRXVRX
FVRXV	TMVRXV	TXRMMMX

Received, April 1997