Implicit learning in adult second language acquisition: abstractness and awareness[1]

Anna Leonard Cook
Linguistics and English Language
The University of Edinburgh
A.J.Leonard-Cook@sms.ed.ac.uk

Research into how people learn non-linguistic artificial grammars suggests that participants use rule knowledge, without being aware of it. More recent evidence suggests that they are actually using surface information, allowing them to behave as if they had learnt rules. Therefore they have no awareness of the rules because they do not know them at all. These two interpretations represent two dimensions: awareness and abstractness. Both are considered in this paper.

Participants were exposed to a Persian target structure without any explicit teaching of the grammar and then tested to see what they had acquired. Some (the Learners) were found to develop explicit abstract knowledge. All learnt the grammar implicitly. No evidence was found that participants relied on the surface characteristics of the sentences, either implicitly or explicitly. Possible explanations of the results are discussed.

Introduction

Implicit learning occurs without awareness or understanding of what has been learnt (Schmidt, 1993). The resulting implicit knowledge can be accessed quickly by gut feeling; it is not possible to report implicit knowledge verbally. It is sensing that somebody looks familiar, rather than the recognising her as your colleague. The term acquisition has been used to refer specifically to implicit learning of languages (Krashen, 1982).

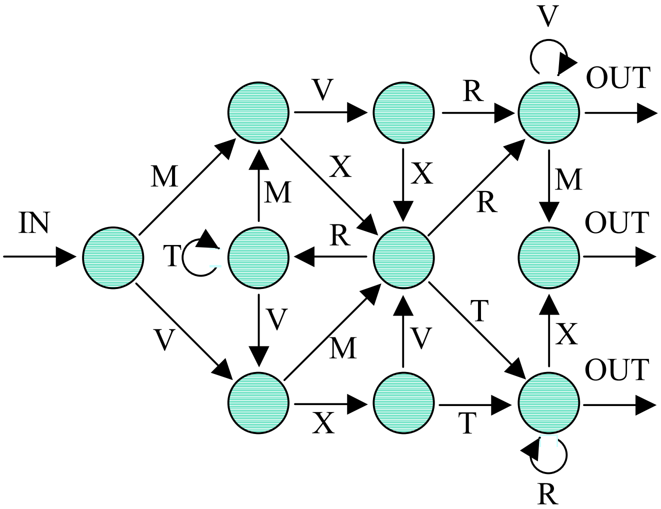
Implicit learning has long been a popular field of research addressing the important question "How can/do human beings learn?" Answering this is vital for maximising learners' potential. It also allows for comparisons between different groups and situations.

Implicit learning research addresses three main issues: a) whether it exists; b) the kind of information that can be learnt implicitly; and c) the relative success of implicit versus explicit learning. Second language acquisition has focused on the third and the evidence has supported explicit learning (Norris and Ortega, 2000). Psychologists have been interested in whether it exists, and what can be learnt implicitly. These are more basic questions – if there is no such thing as implicit learning, explicit learning is bound to be superior.

---

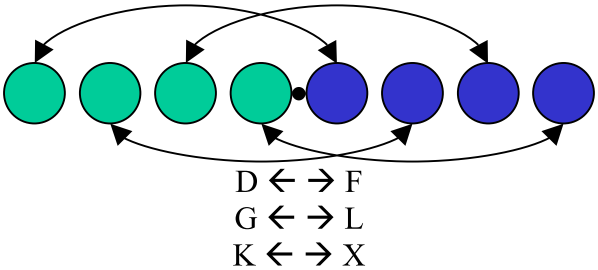
[1] A shorter version of this paper was given at the Moving Forward Postgraduate Conference, College of Arts and Social Sciences, University of Aberdeen

In order to investigate learning in an experimental setting, participants must be given something to learn. Psychologists have used artificial finite-state grammars, an example of which is shown in Figure 1. It generates strings of letters rather than sentences. Start at the in-arrow at the left, and follow the arrows to generate a path to an out-arrow. For example, you could go diagonally upwards (collecting the letter M), right (collecting V), down (collecting X), diagonally down to the right (collecting T), round the loop (collecting R) and finally out. In this case, the total string would be MVXTR. Any string of letters which can be built following these rules is grammatical, and any which cannot is ungrammatical. Participants are exposed to some of the grammatical strings without seeing the grammar itself and asked to memorise them. Next, they have to decide whether a series of new strings follow the rules of the grammar or not. Participants can discriminate grammatical from ungrammatical strings with an accuracy rate of approximately 70%. They cannot explain how they do it nor report any rules that they are using, leading to the conclusion that they lack awareness. These frequently replicated findings suggest that there is such a thing as implicit learning (Reber, 1989).



Looking at Figure 1, it is clear that some pairs of letters (known as bigrams) appear more frequently than others. For example, MV would appear many times, whereas MM never does. Maybe participants do not actually learn the grammar at all, but rather the legal bigrams. They could act as if they knew the rules, as ungrammatical sentences are automatically more likely to contain illegal bigrams. They would be unable to report rules they did not know, and so it would appear that they knew the grammar implicitly (Johnstone and Shanks, 1999).

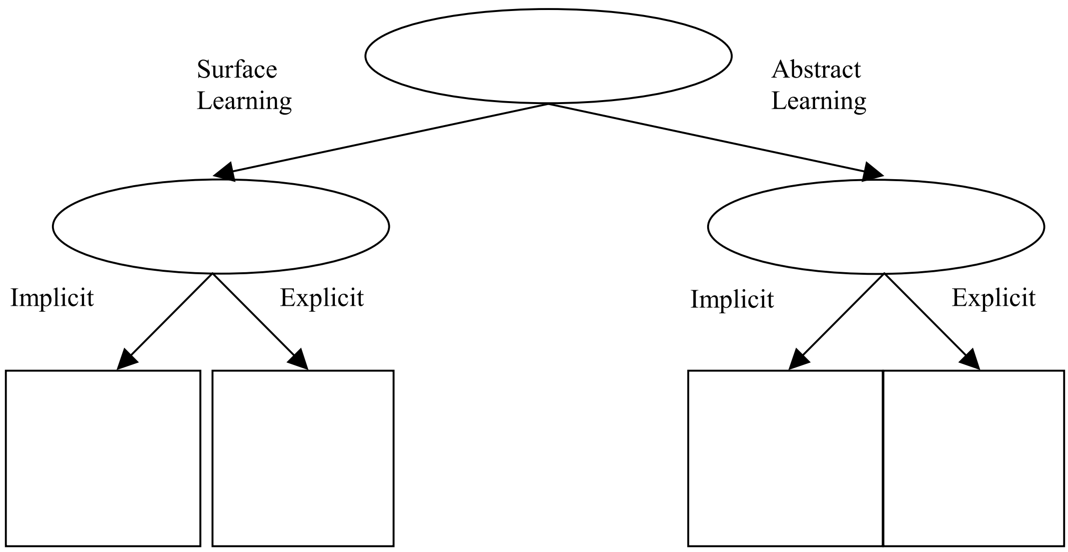
In transitional grammars each letter is determined by the preceding one, and as such it is impossible to separate grammaticality and bigram frequency. Therefore a new type of biconditional artificial grammar was developed which can be seen in Figure 2. These grammars also produce letter strings, in this case using D, F, G, L, K and X. Rather than the preceding letter determining the next, each letter is determined by the one four positions earlier. If the first letter is a D, the fifth must be an F; if the third letter is an L, the seventh must be a G and so on. As such, the grammaticality of a string and the



frequency of its bigrams are unrelated.

Evidence obtained using biconditional grammars supports the interpretation that participants only learn bigram information (Johnstone and Shanks, 2001). Participants believe a new string is grammatical if and only if it contains letter pairs they have seen frequently before, regardless of its grammaticality.

Language provides an ideal testing ground for these theories as it naturally contains long-distance dependencies. This paper reports an experiment in which participants were exposed to a Persian target structure. Afterwards they were tested to see whether they knew word pairs (the equivalent of bigrams) or the grammatical rule. They were assessed to see whether their knowledge was implicit or explicit. Theoretically there were four possible outcomes, as shown in Figure 3.



Method

Participants

24 paid participants were recruited from the University of Edinburgh. None had any prior knowledge of Persian. They had previously learnt at least one foreign language, but had not persevered beyond Standard Grade/G.C.S.E. level or equivalent. In addition, 24 control participants were recruited from the same population.

Procedure

The experiment ran on a computer using E-Prime software. Participants used a button box to answer questions, and occasionally were asked to click the mouse to prevent the screen saver from activating.

The experiment consisted of a learning phase, lasting between 50 and 60 minutes, and a test phase, lasting between 10 and 20 minutes. Afterwards, participants answered a questionnaire about their previous language learning history.

Learning Phase

During the learning phase participants were exposed to Persian sentences. Each trial had the following structure. First, participants saw a picture representing a speaker. Next, a sentence was displayed on the screen for 4.7 seconds, while a recording of it was played. Finally, an English translation appeared for 1.5 seconds. The durations were determined after piloting the procedure.

After four such trials, participants answered a question to ensure that they were following instructions and paying attention to the sentences. A picture of a speaker appeared on the screen, and was followed by a sentence. The participants were asked whether the speaker had said that sentence in the previous four trials. They had as long as they wished to respond.

Prior to the experimental trials, participants were given the chance to practice the procedure with eight trials (and therefore two questions) in German. After this, there were eight trials (and therefore two questions) using single Persian words. Finally, the trials with Persian sentences began. Each sentence was presented eight times during the course of the experiment. The order of presentation was randomised for each participant.

Test Phase

The test phase was taken immediately after the learning phase had been completed. There were three sections, always in the same order.

In the timed grammaticality judgement test participants listened to forty target items and sixteen fillers. Using audio stimuli allowed the exposure times to be controlled naturally (people do not all read at the same speed, but they must listen at the rate at which the stimulus is played). In the ungrammatical target items, the error did not become apparent until the end. Participants then had two seconds, during which the screen changed colour, to indicate whether or not they believed the item was correct[2].

The untimed grammaticality judgement test followed a similar procedure. The stimuli were written instead of spoken, and they remained on screen until a response was logged. This gave participants as long as they wished to examine the sentences before answering. A different set of forty target items and sixteen fillers was used.

Finally there was a multiple-choice sentence-correction task in which participants were shown twelve ungrammatical sentences (four of which were target items). In each case, four corrections were suggested. Participants had to select which would give a completely correct sentence.

Materials

Structure and vocabulary

The target structure was subject-verb agreement for number in Persian intransitive sentences[3]. Only possessive sentences were used, so that each item contained two nouns. As can be seen in 1, this created a long-distance dependency.

---

[2] Piloting had shown that two seconds were required for a sufficient proportion of responses to be logged.

[3] There is an additional accusative particle in transitive sentences which intervenes between the object and the verb, and therefore made them unsuitable.

1)		Possessee-(plural)-of	possessor(-plural)	verb-agreement
	Eg.	*Shohar-an-e*	*yateem*	*tars-ad*
	Eg.	Cousin-s-of	doctor	arrived-plural
	Eg.	The cousins of	the doctor	arrived

Some modifications were made to the form and meaning of the lexical items and suffixes, but not to the structure itself. The nouns were controlled so that they all had the same number of letters (6) after transliteration into the Latin alphabet, the same number of phonemes (5) and the same number of syllables (2). The verbs were subject to the same restrictions, with the exception that *tars* contains four phonemes, whereas the others have only three. Finally, the verb endings were changed. The third person singular has null agreement in the past tense, so the present tense colloquial ending of *–e* was used instead. The plural ending of *–and* was reduced to *–ad* to reduce its similarity to the nominal plural marker *–an*. The full vocabulary used in the experiment can be seen in 2.

2)	**Nouns**	**Meaning in experiment**
	Mehmun	Guest
	Shohar	Cousin
	Yateem	Doctor
	-e	Suffix: of
	-an	Suffix: plural
	Verbs	**Meaning in experiment**
	Geer	Ran
	Kesh	Left
	Tars	Arrived
	-e	Suffix: singular
	-ad	Suffix: plural

The fillers in the learning phase were transitive sentences, using the extra vocabulary in 3, and following the structure shown in 4.

3)	*Dozd*	Saw
	Ro	Accusative particle

4)	Subject	Object	Object Particle	Verb
	Mehmu	*Yateem*	*Ro*	*Dozde*
	Guest	Doctor		Saw
	The guest saw the doctor			

Learning items

Naturally, all of the learning items were grammatical, and so had matching agreement in the first noun and the verb. In addition, when the verb was *geer* or *tars* the intervening second noun always had the same agreement[4]. In consequence, participants were not exposed to noun-verb sequences with non-matching agreement. These sequences were used in the test phase, to create unfamiliar sentences.

[4] (Breaking the pattern for items with kesh gave participants the opportunity to learn that it was the first noun that controlled the agreement).

Test items

There were three independent variables in the test phase as in 5, each with two levels.

5)	Grammaticality	Grammatical	Ungrammatical	Repeated measures
	Familiarity	Familiar	Unfamiliar items	Repeated measures
	Condition	Trained	Control	Independent measures

In grammatical test items the first noun and the verb agreed for number; in ungrammatical items they did not. The second noun could have either singular or plural agreement.

In unfamiliar test items the second two-word sequence (the second noun and the verb) had not been seen in the learning phase because of mismatching agreement. In some cases the lexical items had not been seen together either, regardless of their agreement markers. Table 1 shows how Grammaticality and Familiarity interacted.

		GRAMMATICAL		UNGRAMMATICAL	
		Sing-?-Sing	**Pl-?-Pl**	**Pl-?-Sing**	**Sing-?-Pl**
FAMILIAR	**?-Singular-Singular**	Sing-Sing-Sing		Pl-Sing-Sing	
	?-Plural-Plural		Pl-Pl-Pl		Sing-Pl-Pl
UNFAMILIAR	**?-Plural-Singular**	Sing-Pl-Sing		Pl-Pl-Sing	
	?-Singular-Plural		Pl-Sing-Pl		Sing-Sing-Pl

Audio stimuli

The spoken sentences (lasting three seconds) were built from recordings of individual words. The Persian recordings were made by a native speaker, and the German ones by the experimenter.

Results

Unless otherwise stated, all ANOVA analyses were subject to a $p < 0.05$ significance level. Subsequent t-tests were one-tailed and were subject to a $p < 0.01$ significance level to reduce the family-wide error rate.

Timed Grammaticality Judgement Test

In the timed grammaticality judgement test there was a statistically significant interaction Grammaticality x Condition F_1 (1,46) = 9.122, $p < 0.005$; F_2 (1,36) = 4.681, $p < 0.05$. As can be seen by comparing Figures 4 and 5, the control and trained participants were affected differently by the grammaticality of a test item. There was no equivalent Familiarity x Condition interaction F_1 (1,46) = 1.769; F_2 (1,36) = 1.109. However the Grammaticality effect showed that some knowledge was gained during the learning phase. The responses of the trained participants were then analysed in more detail to see exactly what was learnt.

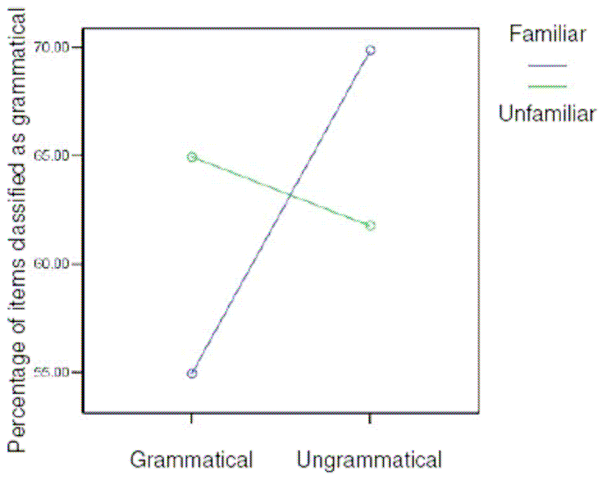
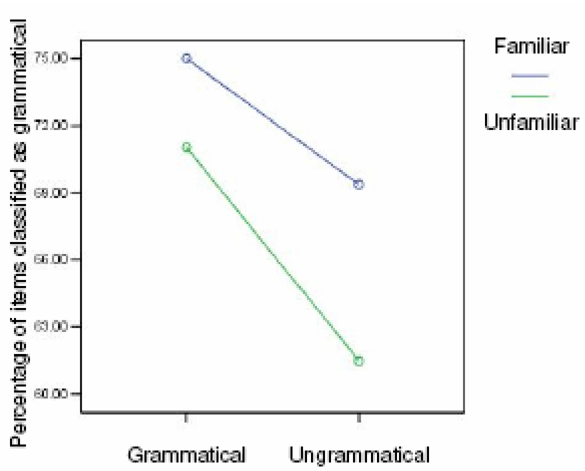


Figure 4: Control participants' responses to the timed grammaticality judgement test



For trained participants there was a significant main effect of Grammaticality $F_1(1,23) = 5.52$, $p < 0.05$; $F_2\ (1,36) = 6.445$, $p < 0.05$. Participants were marginally more likely to accept G-NF items than NG-NF items in a by-subjects analysis $t_1(23) = 1.967$, $p < 0.05$, and significantly so by items $t_2\ (26) = 2.657$, $p < 0.01$. There was no such statistical difference between G-F and NG-F items $t_1(23) = 1.268$; $t_2(10) = 1.379$. Generally trained participants utilised abstract knowledge of the grammar in the timed test, particularly with unfamiliar stimuli.

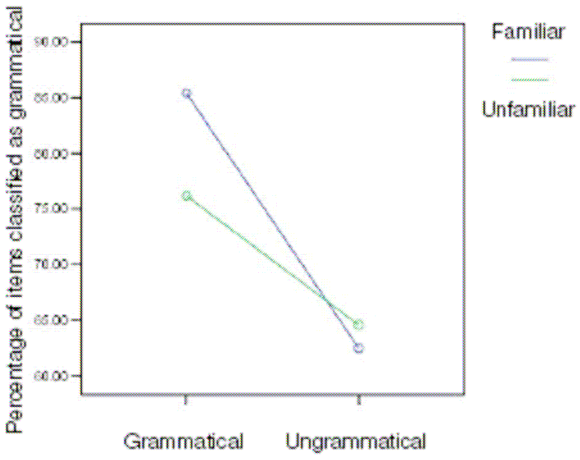
There was only a marginal main effect of Familiarity $F_1(1,23) = 3.981$, $F_2\ (1,36) = 2.794$. T-tests showed that it was not significant for grammatical items $t_1(23) = 0.827$; $t_2(18) = 0.733$ and marginal for ungrammatical ones $t_1(23) = 1.906$, $p < 0.0345$; $t_2(18) = 1.833$, $p < 0.083$.

In summary, participants developed abstract knowledge in the learning phase. As they were able to apply it under time constraints, I suggest that it was implicit.

Untimed Grammaticality Judgement Test

The trained participants were first compared to the control participants, to see if the former had learnt something useful in the learning phase. Again there was a significant interaction of Grammaticality x Condition $F_1\ (1,46) = 8.187$, $p < 0.01$; $F_2\ (1,36) = 21.534$, $p < 0.001$, suggesting that grammatical information had been learnt. There was no interaction Familiarity x Condition by subjects $F_1\ (1,46) = 1.167$ and a marginal one by items $F_2\ (1,36) = 4.074$, $p < 0.051$.

Looking only at the trained participants whose results are shown in Figure 6, there was a statistically significant effect of Grammaticality $F_1(1,23) = 9.182$, $p < 0.01$; $F_2\ (1,36) = 34.123$, $p < 0.001$. G-F items were accepted significantly more often than NG-F items $t_1(23) = 2.979$, $p < 0.005$; $t_2(10) = 4.044$, $p < 0.0025$. G-NF items were accepted marginally more often than NG-NF items in a by-subjects analysis $t_1(23) =$



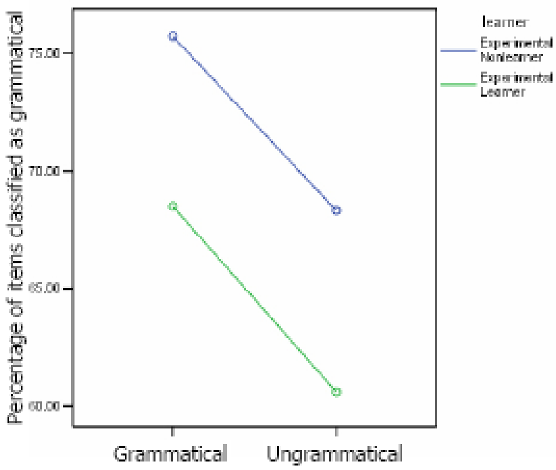
2.221, $p < 0.025$ and significantly more often in a by-items analysis t_2 (26) = 4.651, $p < 0.0005$. Overall, the grammaticality of a sentence affected whether trained participants accepted it as correct or not, regardless of its familiarity.

There was no significant main effect of Familiarity, although it was marginal in a by-items analysis $F_1(1,23) = 0.561$, F_2 (1,36) = 3.791. Participants were marginally more likely to call a G-F sentence grammatical than a G-NF sentence $t_1(23) = 2.104$, $p < 0.025$; t_2 (18) = 2.025; $p < 0.029$. The marginal effect of familiarity did not carry over into ungrammatical sentences. As can be seen in Figure 6 participants were actually more likely to call an NG-NF sentence grammatical than an NG-F item. Therefore the marginal Familiarity effect can be reduced to a benefit for G-F sentences. As these were repeated from the learning phase, this can be explained with whole-item memory without recourse to word pair/bigram familiarity.

There was evidence of abstract knowledge and whole-item knowledge in the untimed grammaticality judgement test, and no evidence that participants knew fragments of the surface strings. It was initially assumed that this knowledge was explicit.

Learners versus Non-Learners

The multiple-choice sentence-correction task was used to divide the trained participants into two groups. The nine Learners, who had successfully corrected three out of four of the sentences, were judged to have explicit knowledge. Only the suggested changes were presented rather than the completed sentences, so it could not have been achieved by implicit knowledge. The fifteen Non-Learners managed less. Learner was treated as a new independent variable, with two levels.



As shown in Figure 7 there was no interaction between the variables Learner and Grammaticality in the timed test F_1 (1,22) = 0.005; F_2 (1,36) = 0.018, nor was there between Learner and Familiarity F_1 (1,22) = 0.85; F_2 (1,36) = 0.388. Learners and Non-Learners used abstract knowledge to the same extent and neither relied on surface knowledge. As the Learners' explicit knowledge could not help them achieve better results, the assumption that the timed test would only be sensitive to implicit knowledge was supported. The Learners and the Non-Learners both developed abstract implicit knowledge.

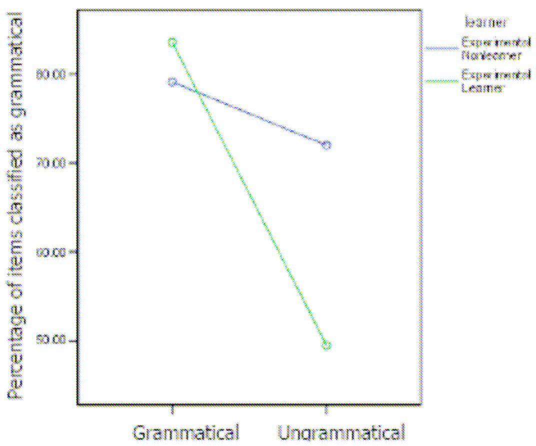


Figure 8 shows that there was a significant interaction between Learner and Grammaticality in the untimed test F_1 (1,22) = 6.521, $p < 0.05$; F_2 (1,36) = 24.953, $p < 0.001$. Learners' explicit knowledge of the grammar made them more sensitive to the grammaticality of a test item than the Non-Learners were. There was no interaction between Learner and Familiarity F_1 (1,22) = 0.433, F_2 (1,36) = 3.459. The Learners and Non-Learners used surface-based information in the same way: not at all. Where it existed, explicit abstract knowledge was used in the untimed test.

Summary

In conclusion, implicit abstract knowledge was used in the timed grammaticality judgement test. In addition, the Learners were able to use explicit abstract knowledge in the untimed test.

Discussion and Conclusion

Participants learnt the abstract rules of the Persian grammar, both implicitly and explicitly. This is in direct contrast with the previous findings reported above, where participants learn surface features (Johnstone and Shanks, 1999). How can this be explained?

Successful inductive explicit learning requires a relatively simple target structure (DeKeyser, 1995), otherwise problem-solving strategies will often lead the learner in the wrong direction. The participants in this study had a distinct advantage in this regard. They were aware that the learning target was a language and therefore rule-governed. They had encountered at least one foreign language in the past, giving them explicit knowledge of the type of structures it can contain, including subject-verb agreement, the target here. Thus, the appearance of explicit abstract knowledge in this experiment is not surprising.

It is not immediately apparent why implicit abstract learning occurred in this experiment, unlike previous research. It may be that learning a language (instead of something else) affects implicit learning as well as explicit learning. This could be because of the syntactic structure, or the included semantics or phonology. In future I will address these issues. Firstly, I will use words rather than individual letters in an artificial grammar, so that the output looks like a language (and participants believe that it is) when structurally it is not. Secondly, I will use individual letters instead of words in a linguistic grammar, so that participants believe they are not dealing with a language when actually they are.

A procedural difference could also have been responsible for the appearance of implicit abstract learning here. This experiment and previous artificial grammar research both used a learning phase during which participants completed a memorisation task. In this experiment, the memorisation task was source-localisation. Participants were asked to match sentences with the speakers who had said them. Traditionally, reproduction tasks were used, whereby participants reproduced the letter strings from memory. It is certainly plausible that reproduction requires greater attention to surface detail than does source-localisation. The next experiment shall investigate this possibility.

This paper described experimental evidence that, at least in some circumstances, abstract grammatical rules can be learnt implicitly. Whether or not this occurs is likely to depend on the nature of the rules and also on the way in which the input is processed. When the situation is suitable for implicit learning to occur, it does so regardless of whether the material is also mastered explicitly.

References

Brooks, L. R. and Vokey, J. R. (1991). Abstract analogies and abstracted grammars: comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General, 120:3,* 316-323.

DeKeyser. R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17*, 379-410.

Johnstone, T. and Shanks, D. R. (1999). Two Mechanisms in Implicit Artificial Grammar Learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory and Cognition, 25:2*, 524-531.

Johnstone, T. and Shanks, D. R. (2001). Abstractionist and Processing Accounts of Implicit Learning. *Cognitive Psychology, 42*, 61-112.

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.

Mathews, R. C., Buss, R. R, Stanley, W. B., Blanchard-Fields, F., Cho, J. R. and Druhan, B. (1989). Role of Implicit and Explicit Processes in Learning From Examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15:6*, 1083-1100.

Norris, J. M. & Ortega, L. (2000). Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis. *Language Learning, 50,* 417-528.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*, 219-235.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13*, 206-226.