available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

# Instruction effects in implicit artificial grammar learning: A preference for grammaticality

Christian Forkstam[a,b,c,*], Åsa Elwér[a], Martin Ingvar[a], Karl Magnus Petersson[a,b,c,d]

[a]Cognitive Neurophysiology Research Group, Stockholm Brain Institute, Karolinska Institutet, Stockholm, Sweden
[b]F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands
[c]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[d]Cognitive Neuroscience Research Group, Universidade do Algarve, Faro, Portugal

ARTICLE INFO

*Article history:*
Accepted 5 May 2008
Available online 13 May 2008

*Keywords:*
Artificial grammar
Implicit learning
Preference classification
Structural mere exposure

ABSTRACT

Human implicit learning can be investigated with implicit artificial grammar learning, a paradigm that has been proposed as a simple model for aspects of natural language acquisition. In the present study we compared the typical yes–no grammaticality classification, with yes–no preference classification. In the case of preference instruction no reference to the underlying generative mechanism (i.e., grammar) is needed and the subjects are therefore completely uninformed about an underlying structure in the acquisition material. In experiment 1, subjects engaged in a short-term memory task using only grammatical strings without performance feedback for 5 days. As a result of the 5 acquisition days, classification performance was independent of instruction type and both the preference and the grammaticality group acquired relevant knowledge of the underlying generative mechanism to a similar degree. Changing the grammatical stings to random strings in the acquisition material (experiment 2) resulted in classification being driven by local substring familiarity. Contrasting repeated vs. non-repeated preference classification (experiment 3) showed that the effect of local substring familiarity decreases with repeated classification. This was not the case for repeated grammaticality classifications. We conclude that classification performance is largely independent of instruction type and that forced-choice preference classification is equivalent to the typical grammaticality classification.



## 1. Introduction

Humans are equipped with acquisition mechanisms with the capacity to implicitly extract structural regularities from experience (Reber, 1967; Stadler and Frensch, 1998). Reber (1967) showed that humans can classify strings generated from an implicitly acquired artificial grammar and he suggested that this process is intrinsic to natural language learning. Implicit learning has four characteristics: (1) no or limited explicit access to the acquired knowledge; (2) the acquired knowledge is more complex than simple associations or exemplar-specific frequency-counts; (3) is an incidental consequence of information processing; and (4) does not rely on declarative memory (Forkstam and Petersson, 2005; Seger, 1994). In this context, we note that several studies (see e.g. Bahlmann et al., 2006; Fletcher et al., 1999) use stimulus material generated from artificial grammars in combination with explicit problem solving tasks with feedback (for a brief review see Petersson

* *Corresponding author*. Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. Fax: +31 24 3521213.
E-mail address: christian.forkstam@mpi.nl (C. Forkstam).

0006-8993/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2008.05.005

et al., 2004). It is important to distinguish these explicit experiments from the type of implicit learning experiments outlined below.

Natural language acquisition is a largely spontaneous, non-supervised, and self-organized process. The structural aspects of natural language are acquired at an early age and largely without explicit feedback (Chomsky, 1965; Jackendoff, 2002; Pinker, 1994; for a different view see e.g. Goldstein and Wetherby, 1984; Hirsh-Pasek et al., 1984; Moerk, 1980). In contrast, reading and writing are examples of typically explicitly taught cognitive skills (see e.g. Petersson et al., in press). Recently, the artificial grammar learning (AGL) paradigm has been proposed as a model for aspects of language acquisition (Gomez and Gerken, 2000; Petersson et al., 2004) and for exploring differences between human and animal learning relevant to the faculty of language (Hauser et al., 2002). Evidence from functional neuroimaging data is consistent with this suggestion. Brain regions related to natural language syntax are also engaged in artificial syntactic processing. In particular, the left inferior prefrontal cortex centered on Broca's region (Brodmann's area, BA, 44/45) is sensitive to artificial syntactic violations (Forkstam et al., 2006; Petersson et al., 2004). Moreover, this region is specifically sensitive to the structural properties rather than to local linear surface features of the input items.

The artificial grammar learning paradigm is a suitable model for the structural aspects of language acquisition. The underlying grammar supports unbounded parsing and generation, and the paradigm comprise implicit learning on acquisition sets of grammatical examples alone without performance feedback (Forkstam et al., 2006; Petersson et al., 2004). It is likely that natural and artificial language acquisition share implicit acquisition mechanisms, as originally suggested by Reber (1967). Additional support for the implicit character of artificial grammar learning comes from lesion studies on amnesic patients. Knowlton and Squire (1996) investigated artificial grammar learning in amnesic patients and normal controls on grammaticality classification. Both groups performed similarly on grammaticality classification, while the amnesic patients showed no explicit recollection of either whole-item or substring information, suggesting that artificial grammar learning depends on the implicit acquisition of structural knowledge (i.e., "rule-based" representations). Alternative theoretical frameworks have questioned the abstract ("rule") acquisition picture and suggest instead that grammaticality classification utilizes exemplar-based representations (Vokey and Brooks, 1992) or substring representations (Perruchet and Pacteau, 1991). In order to address this issue and to control as well as test for any potential substring dependency, the ACS measure was developed (Knowlton and Squire, 1996; Meulemans and Van der Linden, 1997). Associative chunk strength (ACS) is a statistical measure of the associative familiarity of local substrings (e.g., bi- and trigrams) between a classification item and the acquisition set. It is quantified in terms of the frequency with which its substrings occur in the acquisition set. In this approach, acquired structural and instance specific information is quantified by grammaticality and ACS, respectively. From several studies which control ACS it is clear that structural knowledge is acquired (Forkstam et al., 2006; Meulemans and Van der Linden, 1997). This is also consistent with the fact that long-distance dependencies are implicitly acquired in AGL (e.g., Poletiek, 2002).

Taken together, the evidence suggest that grammar learning — whether natural or artificial — can be conceptualized both in terms of structure based rule acquisition and surface based statistical learning mechanisms and not as typically has been proposed as either one or the other. We have recently proposed an alternative view on AGL somewhere between these two conceptualizations (Forkstam et al., 2006; Petersson et al., 2005). In essence, our proposal re-traces a major trend in theoretical linguistics in which syntax is "shifted" into the mental lexicon and where the distinction between lexical items and grammatical rules is beginning to vanish (Culicover and Jackendoff, 2005; Jackendoff, 2002, 2007). In brief, hierarchically structured information is recursively constructed from primitive structures which are stored in long-term memory. On-line integration of structured information results from the unification or successive merging of primitive structures, which are retrieved from long-term memory to a unification component of working memory when activated. Now, if a mechanism for on-line structural integration is already in place, then there is no need for a specific "rule" acquisition mechanism in order to establish a parsing process. In a sense, "rule" acquisition is accomplished by lexical acquisition of structured representations and their subsequent on-line unification (cf. e.g., Hagoort, 2004; Jackendoff, 2007; Vosse and Kempen, 2000).

The typical artificial grammar learning experiment includes a short acquisition session followed by a classification test. During the acquisition phase, participants are engaged in a short-term memory task using an acquisition sample of symbol sequences generated from an artificial grammar, typically a right-linear phrase structure grammar (e.g., Fig. 1, Davis et al., 1994; Perrin and Pin, 2004). Subsequent to the acquisition session, the subjects are informed that the items were generated according to a complex system of rules, without providing information about the rules, and the subjects have to classify new items as grammatical or non-grammatical guided by their immediate intuitive impression (i.e., guessing based on "gut-feeling"); this instruction type will be called the grammaticality classification (GC) instruction in this paper. The subjects

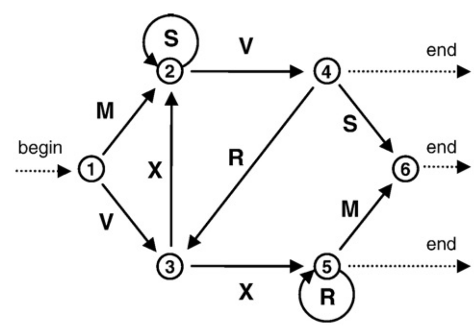


**Fig. 1 – The transition graph representation of the Reber machine used in all experiments to generate the stimulus material. Grammatical strings are generated by traversing the transition graph from state 1 through the internal states along the direction indicated by the arrows (grammatical transitions) until an end state is reached. For example, the grammatical string MSVRXVS can be generated and parsed by the machine through the sequence [1-2-2-4-3-2-4-6] of states, while the non-grammatical string MXVRXSS cannot.**

typically perform reliably above chance, suggesting that they acquired knowledge about relevant aspects of the underlying grammar. Based on the fact that subjects are typically unable to provide sufficient, if any, reasons to motivate their classification decisions (for reviews see Forkstam and Petersson, 2005; Seger, 1994; Stadler and Frensch, 1998), it has been assumed that the classification performance is based on implicit acquisition mechanisms. However, it has been suggested that invoking grammaticality judgments might not be the best way of accessing implicit knowledge, since the grammaticality classification instruction may induce a rule-seeking strategy that, in principle at least, might encourage explicit processing (Manza and Bornstein, 1995; Newell and Bright, 2001). An alternative approach to probe implicit knowledge is based on the mere exposure effect. This effect refers to the finding that repeated exposure to a stimulus induces an increased preference for that stimulus compared to novel stimuli (Zajonc, 1968).

In the present study we investigated the typical artificial grammar learning design while taking advantage of the structural mere exposure effect (Manza and Bornstein, 1995). In mere exposure artificial grammar learning subjects receive preference classification (PC) instruction which make no reference to any previous acquisition episode and the subjects are not informed about the existence of an underlying generative mechanism. The idea is that mere exposure AGL might measure implicit knowledge in a more pure manner since there is nothing in the classification procedures that refers to the acquisition part of the experiment. It has been shown that the preference classification instruction induces similar classification performance as the grammaticality classification instruction (Buchner, 1994; Manza and Bornstein, 1995) in a graded classification task (i.e., preference continuum). This rules out or complicates a direct comparison with forced-choice (yes–no) grammaticality classification. The primary objective of the present study was to compare forced-choice (yes–no) preference with grammaticality classification and to investigate whether and to what extent preference classification would show a similar pattern of results as the standard grammaticality classification. In experiment 1 we directly compared the outcome of implicit artificial grammar learning in subjects given either the grammaticality classification or preference classification instruction. The stimulus material was organized in a 2×2 factorial design using the factors grammaticality (grammatical G/non-grammatical NG) and level of associative chunk strength (ACS; high H/low L). Thus we are able to assess differences between instruction types related to grammaticality as well as of substring familiarity (i.e., ACS). In addition to the rate of acquisition, we also investigated the influence of instruction type on pre-acquisition baseline classification.

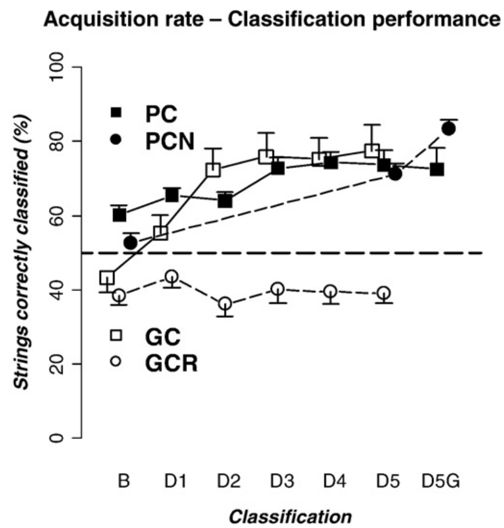


**Fig. 2 – Experiments 1–3: Acquisition rate as a function of classification performance. Strings correctly classified = [number of hits and correct rejections]/[total number of responses] (mean and standard error); B = baseline, D1–5 = day 1–5 classification, D5G = final grammaticality classification, and variable dotted line = response bias deviating from 50% chance level (straight dotted line), see Table 2 for group abbreviations.**

## 2. Results

### 2.1. *Experiment 1 — preference and grammaticality classification*

#### 2.1.1. *Baseline classification*

Instruction type (grammatical/preference classification) influenced the classification behavior already during baseline classification (i.e., pre-acquisition; Fig. 2). Both groups rejected more strings than they accepted, and effects of grammaticality and associative chunk strength developed in the two groups, but in the opposite direction (Fig. 3; preference classification: grammatical > non-grammatical and low > high ACS; grammaticality classification: non-grammatical > grammatical and high > low ACS). To investigate the effect of instruction during baseline, we pooled the participants from experiments 1–3 (i.e., 20 participants with grammaticality classification and 20 with preference classification) and divided the baseline items into two equal sized time-blocks of 20 items (first/second half as they were presented over time). The basic ANOVA was extended with the factor block [1/2]. Any block effect would suggest that the subjects learned properties of the stimulus material already during the baseline classification. The preference classification group increased their rejection rate over time while the rejection rate decreased in the grammaticality classification group, a drift in response bias that differed between groups ($F(1,266) = 4.1$, $P < 0.043$). The effect of grammaticality increased over time in both groups but in the opposite direction. The changing grammaticality effect derives from an increased rejection rate of non-grammatical strings in the preference classification group while this decreased in the grammaticality classification group. The effect of associative chunk strength tended to develop over time in the grammaticality classification group while the preference classification group showed a constant associative chunk strength effect (as well as a tendency to diminish over time, Figs. 3a and b). In summary, the preference classification group acquired sensitivity to grammaticality but showed a constant associative chunk strength effect, while the grammaticality classification group acquired sensitivity to grammaticality and a tendency for an associative chunk

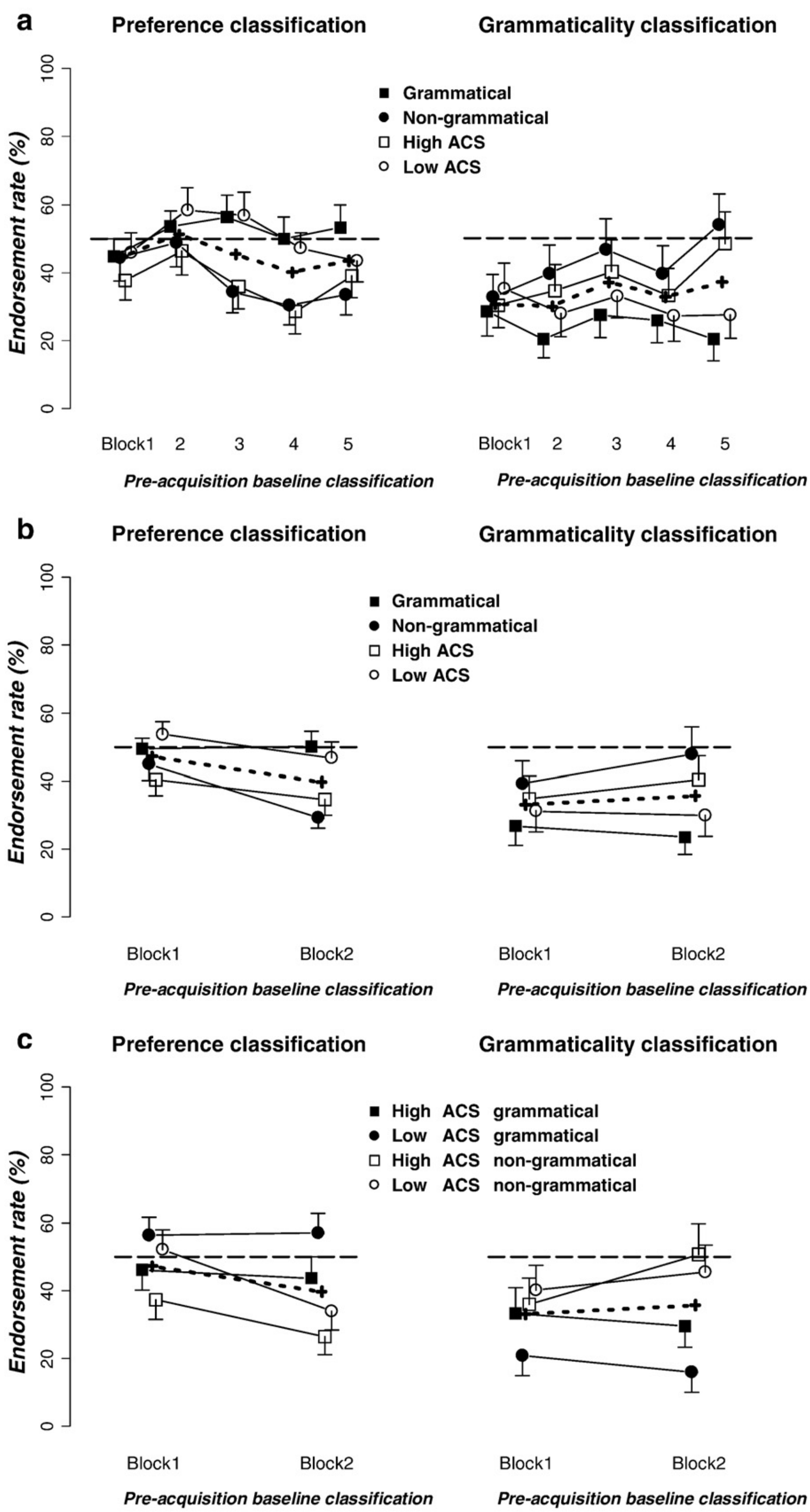


**Fig. 3 – Experiments 1–3: Baseline classification performance. Endorsement rate (mean and standard error) as a function of instruction type: (a) main factors (in 5 blocks); (b) main factors (in 2 blocks); as well as (c) factor levels; Block 1–5 and Block 1–2 = equally sized parts of baseline classification; Dotted line = response bias, ACS = associative chunk strength, and variable dotted line = response bias deviating from 50% chance level (straight dotted line).**

strength effect. However, all effects developed over time, as indicated when dividing the baseline classification performance in more than two blocks (Fig. 3a). The detailed results are outlined in the following two paragraphs.

*2.1.1.1. Effects of grammaticality.* Both the preference and grammaticality classification groups showed effect of grammaticality, although in an instruction dependent manner. The preference classification group rejected non-grammatical strings ($F(1,133)=11$, $P=0.001$), while the grammaticality classification group rejected grammatical strings ($F(1,113)=29$, $P<0.001$; PC vs. GC group: $F(1, 266)=37$, $P<0.001$). This reversed effect of grammaticality increased over time in both groups. The preference classification group rejected non-grammatical strings more often during the second block (block 1: $P>0.24$; block 2: $F(1,57)=12$, $P<0.001$; block 2 vs. block 1: $F(1,133)=3.0$, $P<0.085$), an effect that derived from a significantly increased rejection rate of non-grammatical strings ($F(1,59)=8$, $P=0.008$). In contrast, the rejection rate of non-grammatical strings decreased in the grammaticality classification group ($F(1,59)=6$, $P=0.019$), leading to increased rejection of grammatical strings (block 1: $F(1,57)=6$, $P=0.018$; block 2: $F(1,57)=23$, $P<0.001$; block 2>block 1: $F(1,133)=4.5$, $P=0.035$). Overall, this difference in classification differed between groups ($F(1,266)=7$, $P=0.007$).

*2.1.1.2. Effects of associative chunk strength.* Both groups also showed an effect of associative chunk strength in an instruction dependent manner. The preference classification group preferred low ACS strings and rejected high ACS strings ($F(1,113)=8$, $P=0.004$) while the grammaticality classification group displayed the opposite behavior classifying high ACS strings as grammatical and rejecting low ACS strings ($F(1,113)=3.9$, $P=0.05$; PC vs. GC group: $F(1,266)=12$, $P<0.001$). The associative chunk strength effect did not change significantly over time in either group, but we observed an non-significant increase in the grammaticality classification group (block 1: $F(1,57)=0.8$,

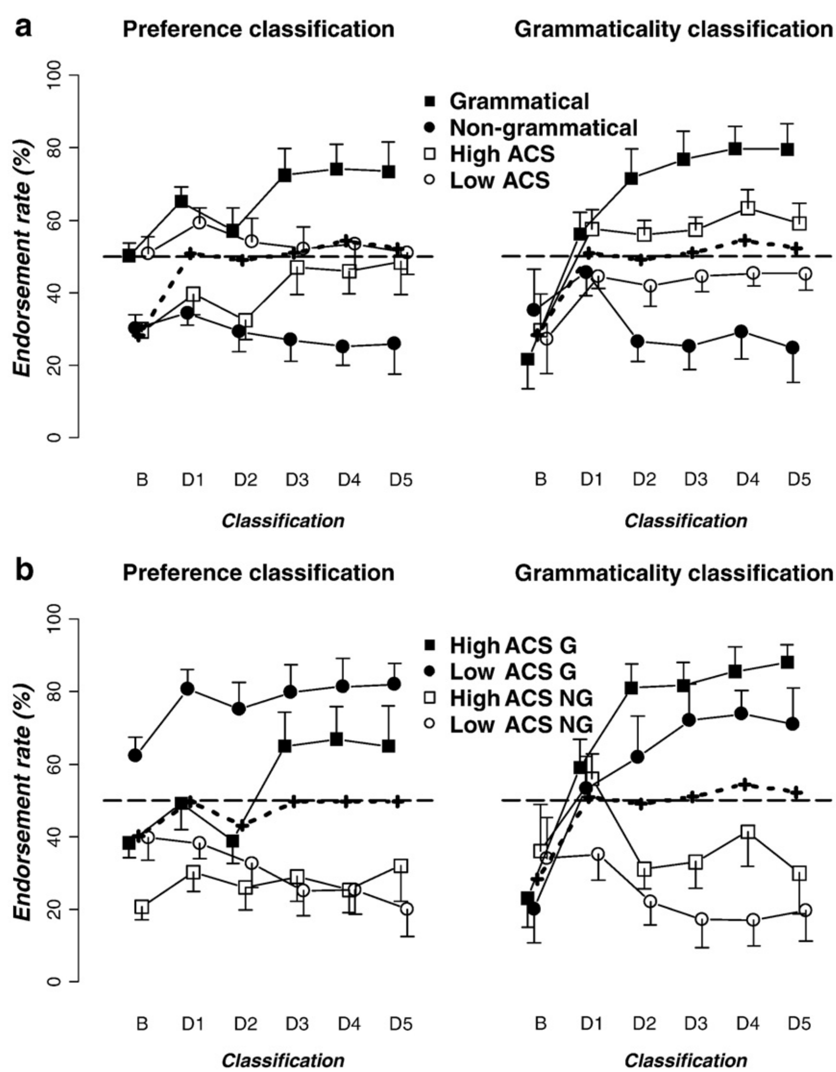


**Fig. 4 – Experiment 1: Classification performance day 1–5. Endorsement rate (mean and standard error) as a function of instruction type: (a) main factors; as well as (b) factor levels; B = baseline, D1–5 = day 1–5 classification, G = grammatical, NG = non-grammatical, ACS = associative chunk strength, and variable dotted line = response bias deviating from 50% chance level (straight dotted line).**

$P>0.36$; block 2: $F(1,57)=3.1$, $P=0.08$; block 2>block 1: $F(1,133)=0.6$, $P>0.4$, ns.), as well as a tendency to decrease in the preference classification group (Fig. 3a).

### 2.1.2. *Classification: day 1–5*

The baseline response bias disappeared immediately after the first acquisition session. Over the 5 days, the grammaticality effect developed in both groups. The grammaticality classification group also switched their preference from non-grammatical to grammatical strings after the first acquisition session, so that the difference between groups with respect to the grammaticality effect disappeared from day 3 onwards. Thus, while both groups acquire sensitivity to grammaticality, the course of acquisition is modulated by instruction type. The effect of associative chunk strength decreased over the 5 days in the preference classification group but stayed constant in the grammaticality classification group. This together with the opposite associative chunk strength effect in the preference and grammaticality classification groups suggest that the associative chunk strength effect is not acquired in the same way as the grammaticality effect. The associative chunk strength effect either does not change over acquisition sessions, as is the case for the grammaticality effect in the grammaticality classification group, or represents only a transient effect that washes out with repeated acquisition, as in the preference classification group (Fig. 4). The detailed results are outlined in the following three paragraphs.

2.1.2.1. *Effects of grammaticality*. Significantly, both groups preferred grammatical over non-grammatical strings during each classification session except for the grammaticality classification group on day 1. However, in the grammaticality classification group the switch in preference from non-grammatical to grammatical strings between baseline and the classification on day 1 was significant ($F(1,63)=4.1$, $P=0.046$). In addition, the grammaticality effect increased in both groups over acquisition sessions, although the increase was larger in the grammaticality than in the preference classification group (day 5>day 1; preference classification: $F(1,63)=4.0$, $P=0.05$; grammaticality classification: $F(1,63)=15$, $P<0.001$; GC>PC group: ($F(1,126)=3.8$, $P=0.05$).

2.1.2.2. *Effects of associative chunk strength*. The preference classification group generally preferred low ACS over high ACS strings, while the grammaticality classification group tended to classify high ACS strings as grammatical more often than low ACS strings. The effect of associative chunk strength decreased over acquisition sessions in the preference classification group (day 1–5: $F(4,171)=2.6$, $P=0.037$; day 1–2: $F(1,27)>13$, $P<0.001$; day 1>day 5: $F(1,63)=4.2$, $P=0.045$), while the effect of associative chunk strength stayed essentially constant over days in the grammaticality classification group (day 1–5: $F(4,171)=0.08$, $P=0.99$, ns.). The grammaticality × associative chunk strength interaction was consistently significant through the acquisition days in the preference classification group ($F(1,171)=23$, $P<0.001$), independent of acquisition session ($F(4,171)=0.4$, $P>0.8$), and it was modulated by instruction type (PC>GC group: $F(1,342)=5.6$, $P=0.018$; GC group: $F(1,171)=0.3$, $P>0.6$).

2.1.2.3. *Effects of structural knowledge and local substring familiarity*. We investigated the effects of grammaticality and associative chunk strength as a function of instruction type and number of acquisition sessions relative to the response bias (i.e., the mean endorsement rate over all four categories). With the assumption that acceptance decisions during classification are based on acquired knowledge about grammaticality or high ACS, and rejection otherwise, two (pure) cases are possible: (1) if structural (i.e., grammatical) knowledge drives classification, the predicted classification pattern over the item types HG/LG/HNG/LNG is accept/accept/reject/reject, while (2) if the classification is driven by local substring familiarity (i.e., ACS) the predicted pattern is accept/reject/accept/reject (cf., Table 1).

Following this logic, the preference classification group acquired structural knowledge already at an early stage, as indicated by their significant preference (relative response bias) for low ACS grammatical strings and rejection of high ACS non-grammatical strings (over all days including baseline; accept LG: baseline: $F(1,19)=7$, $P=0.01$, day 1–5: $F(1,10)>26$, $P<0.001$; reject HNG: baseline: $F(1,19)=11$, $P=0.004$, day 1–5: $F(1,10)>11$, $P<0.007$; Figs. 3c and 4b). Similarly, also the grammaticality classification group acquired structural knowledge. After an initial influence of local substring familiarity during baseline (reject LG: $F(1,19)=12$, $P=0.003$; accept HNG: $F(1,19)=4.8$, $P=0.042$), the grammaticality group switched to accepting low ACS grammatical strings from day 3 and rejecting high ACS non-grammatical strings from day 2 (LG day 1–5: $F(1,36)=4$, $P<0.05$; LG test day interaction term: $F(4,36)=2.5$, $P=0.06$; HNG day 1–5: $F(1,36)=5.4$, $P<0.026$; HNG test day interaction term: $F(4,36)=6$, $P<0.001$; Figs. 3c and 4b). This pattern of results suggests that acquired grammatical knowledge plays a central role in classifying novel strings regardless of instruction type. It provides support for the notion that the basis of preference and grammaticality classification is structural knowledge rather than local substring familiarity.

## 2.2. Experiment 2 — Reber and random strings

In order to establish a reference for the acquisition rate of grammatical structure in the acquisition input, we included

**Table 1 – Structural knowledge vs. local substring familiarity**

| Item category | Structural knowledge (grammaticality) | Local substring familiarity (ACS) |
|---|---|---|
| High ACS–Grammatical | Accept | Accept |
| Low ACS–Grammatical | Accept | Reject |
| High ACS–Non-grammatical | Reject | Accept |
| Low ACS–Non-grammatical | Reject | Reject |

The predicted classification pattern over the item types depending on whether the classification is driven by structural (grammatical) knowledge or local substring familiarity (ACS=associative chunk strength).

an experimental group similar to the grammaticality classification group of the first experiment with the only difference that the Reber strings were replaced by random strings in the acquisition set. Both the grammatical structure group (acquisition on Reber strings) and the random structure group (acquisition on random strings), showed an effect of grammaticality on day 1–5 although in a manner depending on the acquisition set. The grammaticality effect increased over day 1–5 in the grammatical structure group, while this was not the case in the random structure group. There was also an effect of associative chunk strength in both groups over day 1–5. With the same rationale as in experiment 1, using the low ACS grammatical and high ACS non-grammatical dissociation as a marker for structural knowledge vs. local substring familiarity driving the classification performance, the random structure group was guided relatively more by local substring familiarity compared to the grammatical structure group which based their classification decisions primarily on acquired structural knowledge (Fig. 5).

In greater detail, the grammatical structure group generally endorsed grammatical strings while the random structure group endorsed non-grammatical strings (grammatical structure: $F(1,171)=157$, $P<0.001$; random structure: $F(1,171)=62$, $P<0.001$; grammatical vs. random structure: $F(1,342)=218$, $P<0.001$). This effect increased in the grammatical structure group over acquisition days while the endorsement rate of non-grammatical strings remained constant in the random structure group (grammatical structure: $F(4,171)=5.8$, $P<0.001$; grammatical vs. random structure: $F(4,342)=5$, $P<0.001$). Both groups endorsed high ACS compared to low ACS strings to a similar degree, and this effect remained constant over days (grammatical structure: $F(1,171)=18$, $P=0.001$; random structure: $F(1,171)=51$, $P<0.001$). However, the random structure group showed the opposite endorsement behavior with respect

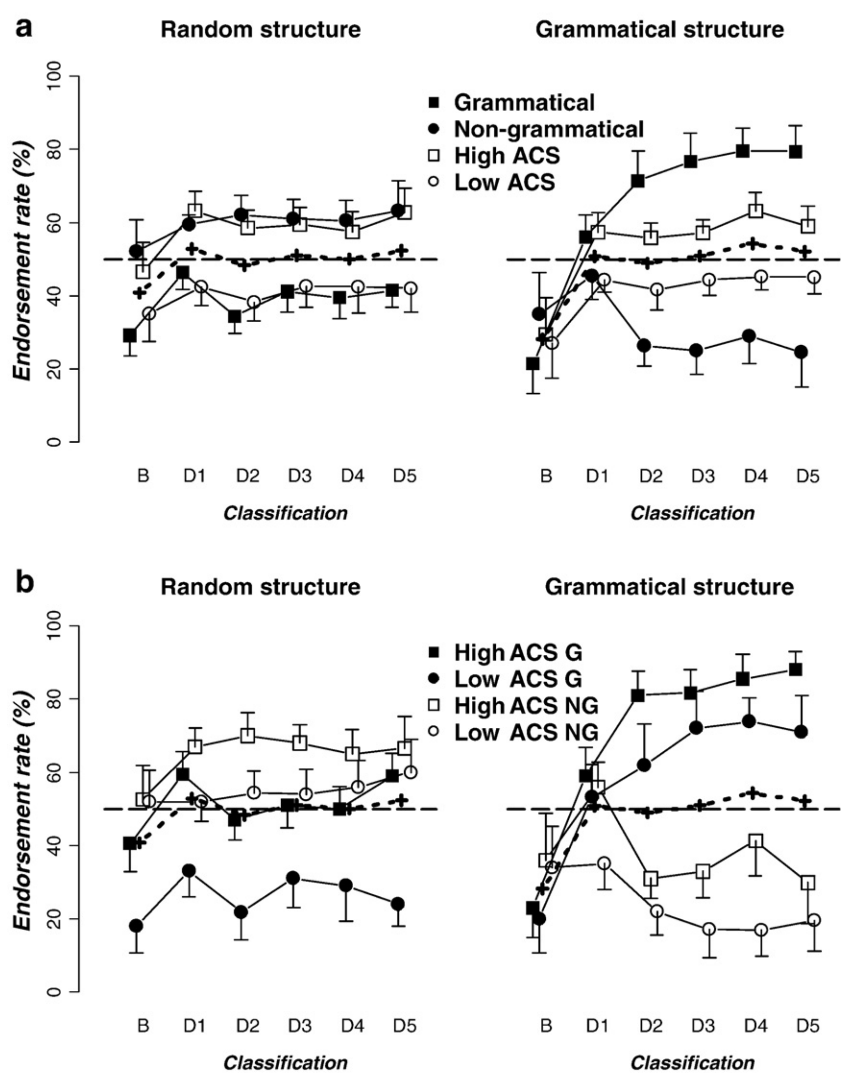


**Fig. 5 – Experiment 2: The effect of acquisition on Reber and random strings. Endorsement rate (mean and standard error) as a function of acquisition material: (a) main factors; as well as (b) factor levels; B = baseline, D1–5 = day 1–5 classification, G = grammatical, NG = non-grammatical, ACS = associative chunk strength, and variable dotted line = response bias deviating from 50% chance level (straight dotted line).**

to low ACS grammatical and high ACS non-grammatical strings compared to the grammatical structure group regardless of instruction type, i.e. compared to both the preference and grammaticality classification groups (LG strings: $F(1,10)>7$, $P<0.03$; HNG strings: $F(1,10)>6$, $P<0.03$; Fig. 5b).

### 2.3. *Experiment 3 — repeated and non-repeated preference classification*

In the third experiment we followed up on the result on preference classification in experiment 1 by investigating the effect of repeated classification. We included an experimental group similar to the repeated preference classification group in experiment 1, with the only difference that they only participated in classification before and after the 5 days of acquisition, that is, non-repeated classification. We found no effect of classification session repetition with respect to grammaticality in either the repeated or non-repeated classification group. Repeated classification did however modulate the effect of associative chunk strength by decreasing the effect of associative chunk strength on endorsement with repeated classification to the degree that the effect was absent from day 3 onwards (Fig. 6). However, on the last day, day 5, the effect of grammaticality increased significantly for the non-repeated classification session group, between the switch from the last preference classification to the final grammaticality classification. The detailed results are outlined in the following two paragraphs.

#### 2.3.1. *Effects of grammaticality*

Both the repeated and the non-repeated classification groups showed an effect of grammaticality to a similar degree on the day 5 final preference classifications (repeated: $F(1,27)=72$, $P<0.001$; non-repeated: $F(1,27)=69$, $P<0.001$; repeated vs. non-repeated: $P>0.4$). This was already the case for the first classifications session (i.e., on day 1 for the repeated group and day 5 for non-repeated group; repeated-day-1: $F(1,27)=33$, $P<0.001$; non-repeated-day-5: $F(1,27)=69$, $P<0.001$; repeated-

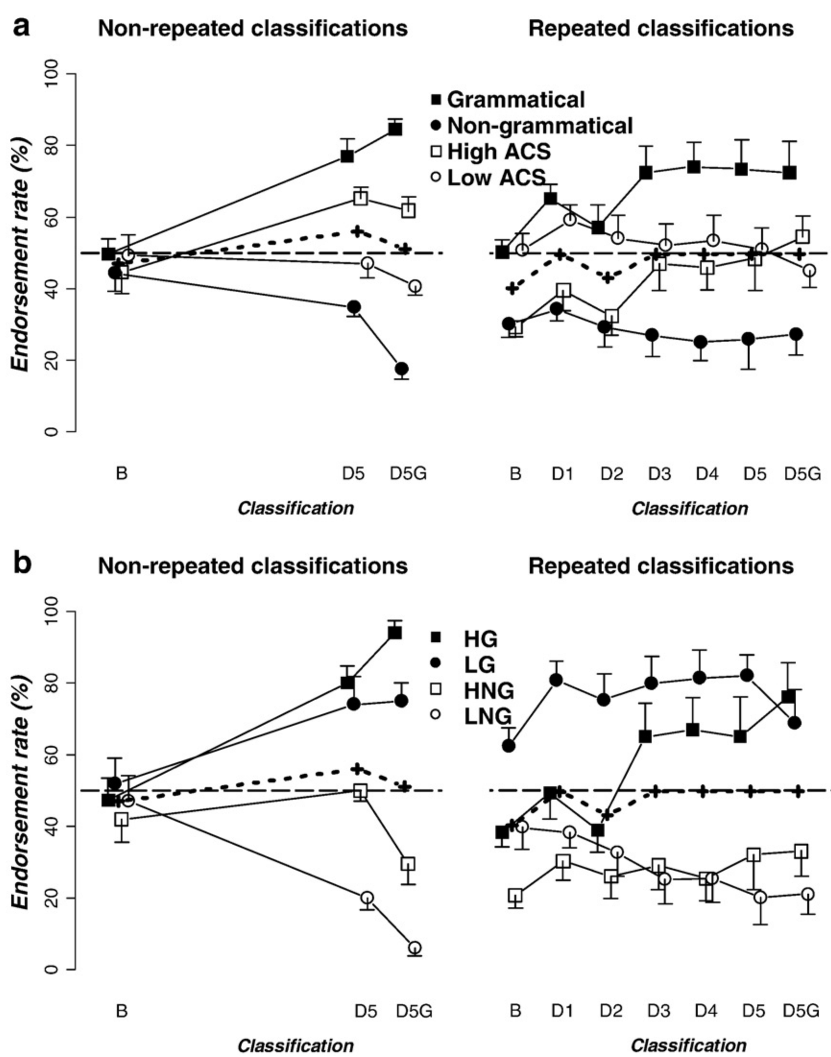


**Fig. 6 – Experiment 3: The effect of repeated classification. Endorsement rate (mean and standard error) as a function of preference classification repetition: (a) main factors; as well as (b) factor levels; B = baseline, D1–5 = day 1–5 classification, D5G = final grammaticality classification, G = grammatical, NG = non-grammatical, ACS = associative chunk strength, and variable dotted line = response bias deviating from 50% chance level (straight dotted line).**

day-1 vs. non-repeated-day-5: $P>0.14$), though in the repeated classification group the effect of grammaticality increased significantly over day 1–5 (day 1: $F(1,27)=72$, $P<0.001$; day 5: $F(1,27)=69$, $P<0.001$; day 5 vs. day 1: $F(1,63)=4$, $P=0.05$). The effect of grammaticality increased significantly in the non-repeated classification group on the final grammaticality classification on day 5, but not in the repeated classification group (non-repeated: $F(1,63)=14$, $P<0.001$; repeated: $P>0.8$; non-repeated vs. repeated: $F(1,126)=4.1$, $P=0.045$).

#### 2.3.2. *Effects of associative chunk strength*

The effect of associative chunk strength differed significantly between groups during the day 5 preference classification, in that the non-repeated classification group preferred high ACS strings while the repeated classification group showed no preference (non-repeated: $F(1,27)=13$, $P=0.001$; repeated: $P>0.6$; non-repeated vs. repeated: $F(1,54)=7$, $P=0.009$). However, the repeated classification group showed an initial effect in the opposite direction for low ACS strings on day 1, which disappeared over days 1–5 (day 1: $F(1,27)=13$, $P=0.001$; day 5: $P>0.6$; day 5 vs. day 1: $F(1,63)=4.2$, $P=0.045$). This pattern of results did not change on the last grammaticality classification (repeated: $P>0.2$; non-repeated: $P>0.6$; Fig. 6).

## 3. Discussion

The main finding of the present study is that preference classification is behaviorally equivalent to the typical grammaticality classification. This is important because the preference version of artificial grammar learning overcomes the potential concern that informing the participants about the existence of an underlying set of rules for generating the acquisition strings, induce explicit strategies. The typical artificial grammar learning setup employs grammaticality classification instruction. Some researchers have raised the concern that this might direct subjects to use explicit problem solving strategies, based on perceived regularities or imagined rules, and that this might interfere with the implicitly acquired knowledge put to use during classification. All experimental groups (except the random-acquisition group of experiment 2) displayed the same qualitative classification behavior overall and acquired knowledge of the underlying syntactic regularities to the same degree. Thus, this concern appears unwarranted, as long as incidental implicit acquisition is employed and the subjects base their classification decisions on their immediate "gut-feeling". Interestingly, the grammaticality effect was boosted in the non-repeated preference classification group when switching from preference to grammaticality instruction (a finding replicated in Folia et al., 2008; Folia et al., in submission). This might suggest that the grammaticality instruction is perceived as more well-defined or focused by the subjects. Alternatively, the grammaticality instruction might trigger general vigilance, motivation, or attention effects. The results also show that the standard grammaticality version of artificial grammar learning assesses implicit acquisition of knowledge about the underlying generative mechanism in the same way as the preference version. However, we suggest that forced-choice preference classification might have certain theoretical advantages over grammaticality classification. It appears that preference classification induces lesser dependency on surface features related to local substrings. Preference classification might thus be less likely to induce explicit rule-based or problem solving strategies, if these latter possibilities are of a real concern. Moreover, the fact that effects of grammaticality as well as associative chunk strength can develop already during classification prior to acquisition suggest that the mechanism engaged (not necessarily the same as in artificial grammar learning proper) can work on surprisingly scarce input. For example, Reber and Perruchet (2003) list a number of features (e.g., number of letters in strings, multiple letter position, letter repetitions, and bigram reoccurrence) which they suggested untrained subjects might employ during baseline or the initial classification phase. However, in this context in the current study, it is important to observe that the participants classified at random at the very beginning of the baseline classification and that the instruction effects developed subsequently over the baseline session (see Fig. 3a).

The first experiment showed that the preference and grammaticality classification groups displayed the same qualitative classification behavior. Both groups acquired structural (grammatical) knowledge to the same degree over the 5 days of acquisition. Thus once the initial transient effects of acquisition had passed, knowledge of the underlying structural regularities was largely independent of instruction type. In other words, the preference classification group started to show preference for grammaticality to the same degree as the grammaticality classification group classified test items correctly. In contrast, the influence of substring familiarity depended on instruction type.

Interestingly, the instruction type modulated the early acquisition pattern during baseline classification. The two types of instruction induced slightly different initial response biases. In addition, the preference classification group developed sensitivity to grammaticality during baseline classification, which continued to develop during the 5 days of acquisition. The grammaticality classification group also showed an effect of grammaticality during baseline classification, but in the opposite direction compared to the preference classification group. However, once the acquisition part of the experiment was initiated, the grammaticality classification group started to endorse grammatical instead of non-grammatical strings, which then developed in the same manner as in the preference classification group. Moreover, both instructions induced effects of substring familiarity but in opposite direction throughout the experiment. Interestingly this effect of substring familiarity washed out as a function of acquisition in the preference classification but not in the grammaticality classification group. Instead, the grammaticality instruction appeared to promote a high ACS string preference that remained throughout the whole acquisition period in the grammaticality classification group.

In the second experiment we establish an acquisition rate reference for the grammaticality classification group. We investigated a group of participants, similar to the grammaticality classification group of the first experiment but with the Reber strings replaced by random strings during acquisition. We found that the classification behavior was driven by local substring familiarity when the acquisition material lacked the underlying grammatical composition (i.e., consisted of random strings). Classification thus predictably did not depend on acquired structural knowledge, as was the case in experiment 1 when grammatical acquisition material was used (Fig. 5).

In the third experiment we followed up on the result on preference classification in experiment 1 by investigating the effect of repeated classification on performance. Repeated classification entails multiple exposures to both positive (grammatical) and negative (non-grammatical) examples. Although no performance feedback was ever given it is conceivable that repeated classification on grammatical and non-grammatical items influence classification performance. It might e.g. change the acquisition dynamics, make the subjects more likely to become aware of the underlying objective of the experiment, or it might make the subjects process the non-grammatical and grammatical information in a non-differentiated manner. We investigated a group of subjects on preference classification before and after the 5 days of acquisition, in contrast to the repeated preference classification in experiment 1. We found that repeated classification had no effect on the acquisition of structural knowledge. In other words, both the repeated and the non-repeated preference classification groups displayed the same classification behavior with respect to grammaticality. On the other hand, the effect of associative chunk strength decreased as a function of classification repetition (Fig. 6). It is thus possible that the increased exposure to negative examples (non-grammatical items) and substrings with low substring familiarity during repeated classification provides the implicit learning mechanism with additional information. Once the processing system has acquired sufficient structural knowledge to distinguish between grammatical and non-grammatical items, this mechanism can make beneficial use of information from negative examples — both non-grammatical and low ACS — in the sense of reducing the influence of associative chunk strength.

In the current series of experiments all groups received their respective classification instruction at the baseline test. Thus the participants were aware what their task would be during the 5 days. This had no qualitative effect on the results with respect to structural acquisition. The same pattern of results is observed in paradigms, which do not include baseline classification, including our own (see e.g. Forkstam et al., 2006; Petersson et al., 2004) and several other AGL experiments (for a review see Forkstam and Petersson, 2005; Pothos, 2007). Thus this aspect of the current paradigm does not result in something new or unexpected. Importantly, the preference groups were never informed about the existence of a complex set of rules generating the acquisition set until their final grammaticality classification test. Nonetheless, preference and grammaticality classification yielded the same result. Moreover, it was emphasized that there was no right or wrong with respect to preference classification and that the subjects should base their classification decision on their immediate gut-feeling. Still preference correlated with grammaticality status independent of local substring familiarity (ACS) on the final day in the preference groups, and to a similar degree as in the grammaticality group. The same holds for the grammaticality group which was also instructed to base their classification decision on their immediate gut-feeling. Although these participants were aware of the existence of a complex set of rules during the 5 days of acquisition, it is all the more surprising that this had virtually no effect on the overall pattern of results, in particular with respect to the acquisition of structural regularities. Again the emphasis was on classification decision based on gut-feeling and the speeded presentation during classification basically precluded any elaborate explicit strategy — this was also what the subjects reported in the post-experiment interview. All subjects complied with the instruction to base their decisions on their immediate gut-feeling and none reported basing their decisions on any substantial rules.

Finally, the non-repeated preference classification group had a final grammaticality classification test on the last day. These participants were never informed about the existence of a complex set of rules for generating the acquisition set until their final grammaticality classification test. This group showed the same overall pattern of results as the grammaticality group and the repeated preference classification group (recently replicated by Folia et al., 2008; Folia et al., in submission). Thus, the conclusion to draw from this set of findings is that the AGL paradigm yields very robust implicit learning quite independent of the experimental details.

## 4. Conclusion

We have shown that classification performance in implicit artificial grammar learning is largely independent of instruction type (preference or grammaticality instruction) and that all experimental groups exposed to grammatical acquisition items acquired structural knowledge related to the underlying generative mechanism to a similar degree. In contrast, surface based substring familiarity was dependent on instruction. This effect decreased with repeated preference classification but not with repeated grammaticality classification. When the underlying grammatical composition of the acquisition material was removed the classification behavior was predictably driven by local substring familiarity and not structural knowledge. We conclude that forced-choice preference classification is equivalent to the typical grammaticality classification. Preference instruction overcomes some concern raised by the fact that instruction on grammaticality depends on informing subjects about the existence of an underlying set of generative rules. We suggest that the preference instruction carry certain theoretical advantages relative the grammaticality instruction in that the classification performance appears to be less dependent on surface features such as substring familiarity, a potential marker for explicit problem solving strategies. On the other hand, our results also show that the standard version of artificial grammar learning assesses implicit acquisition of knowledge about the underlying generative mechanism in the same way as the mere exposure version.

## 5. Experimental procedures

### 5.1. *Experiment 1*

#### 5.1.1. *Participants*

Twenty healthy right-handed university students volunteered to participate in the study (12 females, range 18–40 years).

Participants were pre-screened and none of the subjects used any medication, had a history of drug abuse, neurological or psychiatric illness, or a family history of neurological or psychiatric illness. All subjects were right handed. The experimental protocol was approved by the local Ethics committee, and all participants gave their written informed consent according to the Declaration of Helsinki. To investigate the effect of instruction during baseline, we pooled the participants from experiments 1–3 (i.e., 20 participants with grammaticality classification and 20 with preference classification).

#### 5.1.2. *Experimental groups*

The 20 participants were randomly allocated to 2 groups balanced for gender. All subjects participated in one acquisition session each day for 5 days. The classification instruction was manipulated between groups. Participants in the grammaticality classification (GC) group were informed that they were taking part in an artificial grammar learning experiment consisting of repeated short-term memory experiments (i.e., the acquisition sessions), that a complex set of rules generated the underlying structure in the acquisition material, and they were instructed to classify novel strings as grammatical or not during the classification sessions. It was emphasized that the subjects should base their decisions on their immediate intuition (i.e., guessing based on their "gut-feeling") and avoid any attempt to explicitly analyze strings, since this would yield the best classification performance. Participants in the preference classification (PC) group were not informed about the existence of an underlying generative mechanism until the last classification session which they performed with the grammaticality instruction. They were informed that they were taking part in repeated short-term memory experiments, and that they should classify novel strings as preferable or not (i.e., whether they liked the symbol string or not). Just as for subjects given the grammaticality classification instruction, it was emphasized that they should base their classification decisions on their immediate intuition (i.e., guessing based on their "gut-feeling") and avoid any attempt to explicitly analyze the strings.

The grammaticality instruction was administered to the grammaticality classification group (Table 2) before the baseline (pre-acquisition) classification and each subsequent classification session. Similarly, the preference instruction was administered to the preference classification group before the baseline (pre-acquisition) and each subsequent classification session day 1–5, followed by a final classification session with the grammaticality instruction.

#### 5.1.3. *Stimulus material*

We generated 569 grammatical (G) strings from the Reber grammar (5–12 consonants long from the alphabet {M, S, V, R, X}; see Fig. 1). To derive the associative chunk strength (ACS) for each string, we calculated the frequency distribution of bi- and trigrams (i.e., substrings of length 2 and 3) for both terminal and complete string positions (Knowlton and Squire, 1996; Meulemans and Van der Linden, 1997). In an iterative procedure we randomly selected 100 strings to generate an acquisition set which were representative in terms of associative chunk strength in comparison to the complete string set. For the random acquisition set, 100 random strings were generated from the same alphabet and of the same length and with similar levels of associative chunk strength as the Reber grammar acquisition set. For each remaining grammatical string in the complete string set, we generated non-grammatical (NG) strings by a switch of letters in two non-terminal positions. We selected the non-grammatical string that best matched the grammatical strings in terms of both terminal and complete string position associative chunk strength (i.e., collapsed over order information within strings). These grammatical and non-grammatical strings were further classified in terms of their associative chunk strength status independent of grammatical status. High/low associative chunk strength refers to classification strings composed of common/uncommon bi- and trigrams in the acquisition set (for both terminal and complete string position). In an iterative procedure we randomly selected 6 sets of 40 strings each such that for a given classification set: (1) its high ACS strings did not differ significantly in terms of associative chunk strength compared to the acquisition set; (2) its high and low ACS strings did not differ significantly with the high and low ACS strings in the other classification sets; and (3) its low ACS strings did differ significantly with both the acquisition set and to the high ACS strings in each classification set. Thus the classification material was organized in a 2×2 factorial design with the factors grammaticality (grammatical/non-grammatical) and

**Table 2 – Experiments 1–3: Experimental design**

| Subjects ($N$=10/group) | Baseline | Day 1–4 | | Day 5 | | |
|---|---|---|---|---|---|---|
| | | AQ | CL | AQ | CL | |
| Grammaticality-Random (GCR) | GC | | GC | | GC | |
| Grammaticality (GC) | GC | | GC | | GC | |
| Preference (PC) | PC | | PC | | PC | GC |
| Non-repeated Preference (PCN) | PC | | | | PC | GC |

AQ = acquisition; CL = classification session; GC = grammaticality classification; PC = preference classification; Reber strings = dark grey; Random strings = light grey.

associative chunk strength (high/low), and the classification sets included 10 strings from each category: high ACS grammatical (HG), low ACS grammatical (LG), high ACS non-grammatical (HNG), and low ACS non-grammatical (LNG).

#### 5.1.4. *Procedure*

During acquisition, 100 strings were presented on a computer screen one by one. The string presentation order was randomized for each acquisition session. Each string was presented for 5 s and the participants were asked to type the string from memory after the string disappeared from screen in a self paced manner. Each acquisition task took between 25 and 45 min. The classification sets were balanced across subjects, days and groups, and the string presentation order was randomized for each test. The baseline test was presented to the subjects as independent of the subsequent testing. The preference classification groups were unaware of the final grammaticality classification test until it occurred on the last day, while participants with grammaticality instructions were informed about the existence of an underlying complex set of rules in the acquisition material during the 5 days. Each classification session was presented as a yes/no classification task and lasted approximately 5 min during which 40 strings were presented one at a time for 3 s on a computer screen. The subjects made their classification decision during a 2 s response window.

#### 5.1.5. *Data analysis*

We used a mixed-effect multi-way repeated measures analysis of variance (ANOVA) with endorsement rate (i.e., strings accepted as grammatical/preferable regardless of grammaticality) as the dependent variable using the statistics package R (www.r-project.org). We modeled the main factors grammaticality [G/NG], associative chunk strength [H/L], and classification session [baseline/day 1–5] as within subjects fixed-effects, group [GC/PC] as between subjects fixed-effect, and subjects as random-effects. An overall significance level of $P<0.05$ was used for statistical inference, and explanatory investigations for significant effects were restricted to the reduced ANOVA contrasted over the appropriate factor levels. To investigate the effect of instruction during baseline, we pooled the participants from experiments 1–3 (i.e., 20 participants with grammaticality classification and 20 with preference classification) and divided the baseline items into two equal sized time-blocks of 20 items (first/second half as they were presented over time). The basic ANOVA was extended with the factor block [1/2].

### 5.2. *Experiment 2*

#### 5.2.1. *Participants, experimental groups and procedures*

Ten new healthy right-handed university students volunteered in the study (6 females, range 18–40 years), and participated in a random grammaticality classification (GCR) group (Table 2). For the random grammaticality classification group grammaticality classification instructions were administered during both the baseline (pre-acquisition) classification and during each subsequent classification session. The only difference between the random grammaticality classification and the grammaticality classification group was the nature of the acquisition material in that it consisted of random strings for the random grammaticality classification group while it was derived from the Reber grammar for the grammaticality classification group. In all other respects the participant screening, stimulus material and the experimental procedure was identical to experiment 1.

#### 5.2.2. *Data analysis*

We modeled the main factors grammaticality [G/NG], associative chunk strength [H/L], and classification session [baseline/days1–5] as within subjects fixed-effects, group [GC/GCR] as between subjects fixed-effects, and subjects as random-effects. An overall significance level of $P<0.05$ was used for statistical inference, and explanatory investigations for significant effects were restricted to the reduced ANOVA contrasted over the appropriate factor levels.

### 5.3. *Experiment 3*

#### 5.3.1. *Participants, experimental groups and procedures*

Ten additional healthy right-handed university students volunteered in the study (6 females, range 18–40 years), and participated in a non-repeated preference classification (PCN) group (Table 2). For the PCN group preference instruction were administered during the baseline (pre-acquisition) classification and during the classification session subsequent to the acquisition sessions on test day 5, followed with a final classification session with the grammaticality instruction. The only difference between the non-repeated preference classification and the preference classification group in experiment 3 was the number of classification sessions in that classifications were also administered during each day (1–5) for the preference classification group. In all other respects the stimulus material and the experimental procedure was identical to experiments 1 and 2.

#### 5.3.2. *Data analysis*

We modeled the main factors grammaticality [G/NG], associative chunk strength [H/L], and classification session [baseline/day5PC/day5GC] as within subjects fixed-effects, group [PC/PCN] as between subjects fixed-effects, and subjects as random-effects. An overall significance level of $P<0.05$ was used for statistical inference, and explanatory investigations for significant effects were restricted to the reduced ANOVA contrasted over the appropriate factor levels.

## Appendix — task instructions

### *Short term memory task*

During this task consonant strings will be presented on the computer screen. You shall study the string attentively and after presentation you shall retype the string as correct as possible on the computer keyboard. It is important that you concentrate on each string to remember it correctly.

### *Preference classification*

During this task consonant strings will be presented on the computer screen. You shall study the string attentively during presentation and respond whether you like the string or not based on your immediate impression. Trust your immediate intuitive impression or guessing based on 'gut-feeling' and avoid any other elaboration of the basis for your decision.

### *Grammaticality classification*

The consonant strings presented during the short-term memory test all belong to an artificial language. They were generated according to a complex set of rules, that is, they are all grammatical in relation to these rules. During this task consonant strings will be presented on the computer screen. Half of the set of strings is grammatical and the other half is not. You shall study the string attentively during presentation and respond whether the string is grammatical or not. Trust your immediate intuitive impression or guessing based on 'gut-feeling' and avoid any other elaboration of the basis for your decision. You will achieve the best performance if you base the decisions on your intuitive 'gut-feeling'.

## REFERENCES


Bahlmann, J., et al., 2006. Hierarchical and linear sequence processing: an electrophysiological exploration of two different grammar types. J. Cogn. Neurosci. 18, 1829–1842.

Buchner, A., 1994. Indirect effect of synthetic grammar learning in an identification task. J. Exper. Psychol., Learn., Mem., Cogn. 20, 550–566.

Chomsky, N., 1965. Aspects of the Theory of Syntax. MIT Press, Cambridge, MA. vol.

Culicover, P.W., Jackendoff, R., 2005. Simpler Syntax. Oxford University Press, Oxford, UK. vol.

Davis, M.D., et al., 1994. Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science. Academic Press, San Diego, CA. vol.

Fletcher, P., et al., 1999. Learning-related neuronal responses in prefrontal cortex studied with functional neuroimaging. Cereb. Cortex 9, 168–178.

Folia, V., et al., 2008. The mere exposure effect in artificial grammar learning: implications for dyslexia studies. Proc. Cogn. Neurosci. Master Radboud Univ. 3 (1), 5–19.

Folia, V., et al., in submission. Implicit learning and dyslexia. Annals of the New York Academy of Science.

Forkstam, C., Petersson, K.M., 2005. Towards an explicit account of implicit learning. Curr. Opin. Neurol. 18, 435–441.

Forkstam, C., et al., 2006. Neural correlates of artificial syntactic structure classification. NeuroImage 32, 956–967.

Goldstein, H., Wetherby, B., 1984. Application of a functional perspective on receptive language development to early intervention. RASE: Remedial Spec. Educ. 5, 48–58.

Gomez, R.L., Gerken, L., 2000. Infant artificial language learning and language acquisition. Trends Cogn. Sci. 4, 178–186.

Hagoort, P., 2004. Broca's complex as the unification space for language. In: Cutler, A. (Ed.), Twenty-first Century Psycholinguistics: Four Cornerstones, vol. 20. Erlbaum, pp. s18–s29.

Hauser, M.D., et al., 2002. The faculty of language: what is it, who has it, and how did it evolve? Science 298, 1569–1579.

Hirsh-Pasek, K., et al., 1984. Brown and Hanlon revisited: mothers' sensitivity to ungrammatical forms. J. Child Lang. 11, 81–88.

Jackendoff, R., 2002. Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford University Press, Oxford, UK. vol.

Jackendoff, R., 2007. A parallel architecture perspective on language processing. Brain Res. 1146, 2–22.

Knowlton, B.J., Squire, L.R., 1996. Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. J. Exper. Psychol., Learn., Mem., Cogn. 22, 169–181.

Manza, L., Bornstein, R.F., 1995. Affective discrimination and the implicit learning process. Conscious. Cogn. 4, 399–409.

Meulemans, T., Van der Linden, M., 1997. Associative chunk strength in artificial grammar learning. J. Exper. Psychol., Learn., Mem., Cogn. 23, 1007–1028.

Moerk, E.L., 1980. Relationships between parental input frequencies and children's language acquisition: a reanalysis of Brown's data. J. Child Lang. 7, 105–118.

Newell, B.R., Bright, J.E., 2001. The relationship between the structural mere exposure effect and the implicit learning process. Q. J. Exp. Psychol., A Human Exp. Psychol. 54A, 1087–1104.

Perrin, D., Pin, J.-E., 2004. Infinite Words: Automata, Semigroups, Logic, and Games. Elsevier, Amsterdam, The Netherlands. vol.

Perruchet, P., Pacteau, C., 1991. Implicit acquisition of abstract knowledge about artificial grammar: some methodological and conceptual issues. J. Exp. Psychol. Gen. 120, 112–116.

Petersson, K.M., et al., 2004. Artificial syntactic violations activate Broca's region. Cogn. Sci. 28, 383–407.

Petersson, K.M., et al., 2005. Artificial grammar learning and neural networks. Proc. Cogn. Sci. Soc. 1726–1731.

Petersson, K.M., et al., in press. Language and literacy from a cognitive neuroscience perspective. In: D.R. Olson, N. Torrence, Eds., Cambridge Handbook of Literacy. vol. Cambridge University Press, New York.

Pinker, S., 1994. The Language Instinct. William Morrow & Co, New York, NY. vol.

Poletiek, F.H., 2002. Implicit learning of a recursive rule in an artificial grammar. Acta Psychol. 111, 323–335.

Pothos, E.M., 2007. Theories of artificial grammar learning. Psychol. Bull. 133, 227–244.

Reber, 1967. Implicit learning of artifcial grammars. J. Verbal Learn. Verbal Behav. 6, 855–863.

Reber, R., Perruchet, P., 2003. The use of control groups in artificial grammar learning. Q. J. Exp. Psychol. 56A, 97–115.

Seger, C.A., 1994. Implicit learning. Psychol. Bull. 115, 163–196.

Stadler, M.A., Frensch, P.A. (Eds.), 1998. Handbook of Implicit Learning. Sage Publications, Thousand Oaks, CA.

Vokey, J.R., Brooks, L.R., 1992. Salience of item knowledge in learning artificial grammar. J. Exper. Psychol., Learn., Mem., Cogn. 20, 328–344.

Vosse, T., Kempen, G., 2000. Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. Cognition 75, 105–143.

Zajonc, R.B., 1968. Attitudinal effects of mere exposure. J. Pers. Soc. Psychol. Monogr. Suppl. 9, 1–27.