



**HAL**  
open science

## Implicit learning of two artificial grammars

Camille Guillemin, Barbara Tillmann

► **To cite this version:**

Camille Guillemin, Barbara Tillmann. Implicit learning of two artificial grammars. *Cognitive Processing*, 2021, 10.1007/s10339-020-00996-2 . hal-02999539

**HAL Id: hal-02999539**

**<https://hal.science/hal-02999539>**

Submitted on 10 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Implicit learning of two artificial grammars**

Guillemin, C.<sup>1,2</sup> & Tillmann, B.<sup>1,2</sup>

<sup>1</sup> CNRS, UMR5292; INSERM, U1028; Lyon Neuroscience Research Center, Auditory  
Cognition and Psychoacoustics Team, Lyon, F-69000, France;

<sup>2</sup> University Lyon 1, Villeurbanne, F – 69000, France;

**Short title:** Two artificial grammar learning

Corresponding author:

Barbara Tillmann

Centre de Recherche en Neurosciences de Lyon

Inserm U1028 - CNRS UMR5292 - UCBL

Centre Hospitalier Le Vinatier - Bâtiment 462 - Neurocampus

95 boulevard Pinel

69675 Bron Cedex

France

Email: Barbara.Tillmann@cns.fr

## **Abstract**

This study investigated the implicit learning of two artificial systems. Two finite-state grammars were implemented with the same tone set (leading to short melodies) and played by the same timbre in exposure and test phases. The grammars were presented in separate exposure phases, and potentially acquired knowledge was tested with two experimental tasks: a grammar categorization task (Experiment 1) and a grammatical error detection task (Experiment 2). Results showed that participants were able to categorize new items as belonging to one or the other grammar (Experiment 1) and detect grammatical errors in new sequences of each grammar (Experiment 2). Our findings suggest the capacity of intra-modal learning of regularities in the auditory modality and based on stimuli that share the same perceptual properties.

**Key words:** incidental learning, implicit cognition, artificial grammar, dual-system learning

Implicit or incidental learning is the ability to learn, by mere exposure and without intention, a structured system (e.g., Perruchet, 2008). This cognitive capacity has been shown for verbal, musical or visuo-spatial materials (e.g., Saffran, 2001; Rohrmeier et al., 2011; Fiser & Aslin, 2002), and is relevant for language acquisition, motor learning, and tonal enculturation (e.g., Perruchet & Pacton, 2006; Rebuchat & Rohrmeier, 2012).

Our study investigated whether perceivers can learn two artificial systems at the same time even when implemented with the same events in the same modality. Previous findings are controversial, confirming or disconfirming dual-system learning. They had been obtained with two of the main implicit learning paradigms, notably using artificial grammar (AG) or artificial language (AL), respectively.

In the AG paradigm, participants are exposed to sequences built with a finite-state grammar defining combinations between events (Reber, 1967; see Figure 1). The classical implementation uses letters, but AG learning has been also shown with colours (Conway & Christiansen, 2006), shapes (Emberson et al., 2011), durations (Prince et al., 2018), dance movements (Opacic et al., 2009) and tones (e.g., Altmann et al., 1995; Tillmann & Poulin-Charronnat, 2010). During exposure, participants are not told about the regularities and sometimes perform an additional task requiring memory (i.e., was this sequence or item presented earlier?) or attention (e.g., detecting altered events). In the test phase, participants are informed that all sequences follow rules, and they have to distinguish *new* AG sequences from grammar-violating sequences. Participants' performance above chance (50%) has shown learning of the AG, at least some of its features and regularities.

The AL paradigm has been initially created to investigate the use of statistical cues in word acquisition (e.g., Saffran, Newport, & Aslin, 1996), but has then been applied to numerous other materials, such as tones (Saffran et al., 1999), timbres (Tillmann & McAdams, 2004; Hoch et al., 2010) or visual forms (Fiser & Aslin, 2001; Hunt & Aslin, 2001). Participants are

exposed to a language stream composed of artificial words (mostly of three syllables), which are chained sequentially without pauses or acoustic segmental cues. The syllable stream is either presented as a foreign language or not. The test phase requires participants to discriminate AL words, which occurred in the exposure stream, from non-words (not occurring in the exposure stream) or part-words (only partially or less frequently occurring in the exposure stream). Participants' above chance performance indicated the capacity of AL word learning. Research implementing this paradigm with non-verbal materials has provided evidence for a domain-general statistical learning mechanism (e.g., Saffran et al., 1999; McMullen. & Saffran, 2004)).

In sum, AG and AL research has provided evidence for learning of one single artificial system. However, learning of two artificial systems has been shown only for AL (Franco et al., 2011), but not for AG (Conway & Christiansen, 2006), based also on different exposure phase implementations, respectively (see below). Conway and Christiansen (2006) investigated simultaneous learning of two AGs presented within and across modalities. While participants were able to learn two AGs implemented in different modalities (visual *versus* auditory) or in different materials within the same modality (tones and spoken non-words), they learned only one of the two AGs when these AGs were implemented with the same material (e.g., spoken non-words). As the presentation of the grammatical sequences from the two AGs were randomly intermixed during exposure, the authors acknowledged that the same-material implementation might lead to perceptual confusion during learning or in the acquired knowledge representation. Franco et al. (2011) reduced potential confusion between the two to-be-learned systems by implementing exposure to the two ALs with two subsequent exposure phases and using two voices to pronounce the syllables of the ALs (one for each AL). With this implementation, participants learned the two AL systems (based on the same

syllables, all presented auditorily), successfully distinguishing words of each AL from non-words (which had not occurred during the exposure phase).

The two studies with contrasting conclusions thus differed in particular by the to-be-learned system (AG *versus* AL<sup>1</sup>) and exposure implementation (one exposure phase intermixing stimuli of the two systems (Conway & Christiansen, 2006) versus two exposure phases separating the systems, additionally marked with different voices (Franco et al., 2011)). This latter difference related to the exposure implementation might have more impact on potential dual-system learning (i.e., because of the discussed potential confusion) than the type of structured material to be learned (i.e., AG or AL). In addition, the separated exposure phases imitate somewhat a typical bilingual environment, notably with each language either being presented in temporally connected chunks or periods or being pronounced by a given speaker.

In our present study, we thus used the implementation of two successive exposure phases (one for each to-be-learned system). Further, we decided to use two AGs as their joint learning has not been shown before. We aimed to test for intramodal learning of regularities in stimuli that share the same perceptual properties. For that aim, we implemented the two AGs with the same tone set, which was played by the same timbre for exposure and test. In the experimental session, participants were first exposed to sequences of the two AGs in two separate phases (with counterbalanced order) and then tested with new sequences of both grammars (in a subsequent test phase).

Previous research has suggested that the incidental (implicit) acquisition of grammatical structures is more powerful than the explicit acquisition of the same structures (e.g., Fletcher et al., 2005). Along the same lines, it has been argued that experimental tasks should not reveal links to the grammatical structure (that is, not revealing the rule-governed nature of the

---

<sup>1</sup> In the implementation of the AL paradigm by Franco et al. (2011), non-words did not occur in the exposure stream (i.e., had syllable transitional probabilities of 0).

stimuli to participants or encouraging explicit feature search), but should focus on another aspect of the material, thus promoting indirect investigation methods (e.g., Vinter & Perruchet, 1999; Tillmann & Poulin-Charronnat, 2010). Aiming to benefit from the potential strength of implicit cognition, we did not tell participants about the AG or underlying rules at the exposure phase or at the test phase. For exposure, we insured participants' attention to the material with a tone counting task. At test, participants were told about two composers (one for each exposure phase) and were asked to classify whether each of the new melodies was from the repertoire of composer 1 or composer 2. These two methodological features aimed to maximize learning and test sensitivity to ensure the observation of structure learning, even when two systems were presented. Note that in the previous dual-system learning studies, participants were not told about the artificial systems before exposure, but were informed about the artificial systems at test and judged test items explicitly, notably whether the new stimuli "followed the same rules as before" (Conway & Christiansen, 2006), which of two strings "sounded more like one of the languages" or whether the word was from AL1, AL2 or a new word (Franco et al. 2011).

Finally, in our present study we asked participants to rate their confidence in their response. If learning is implicit, there would be no relation between test phase confidence ratings and accuracy. Dienes et al. (1995) reported that the participants' confidence in incorrect decisions was just as great as their confidence in correct decisions (zero correlation criterion  $\approx 0$ ), suggesting that the participants were unconscious of what they used to make their decision. This "zero correlation criterion" aims to evaluate the unconscious status of what is learned by combining for each participation the information based on accuracy and confidence (Dienes, 2004, 2008).

## **Experiment 1**

### **Method**

**Participants.** Thirty participants took part in Experiment 1. Fifteen participants with a mean age of 21.93 years ( $\pm 1.62$ ) were first exposed to grammar A, then to grammar B, and 15 participants with a mean age of 22.87 ( $\pm 2.42$ ) were first exposed to grammar B, then to grammar A. All participants reported 0 years of musical training (i.e., on instrument, voice or theory).

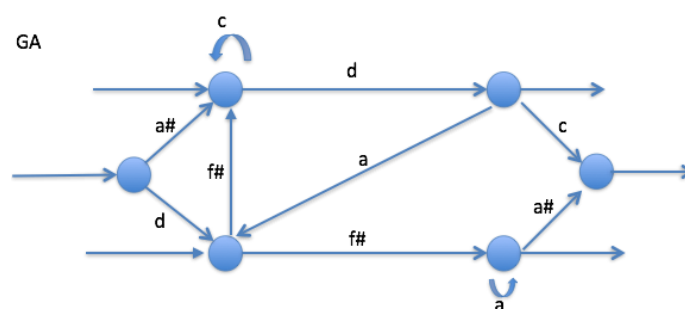
**Material.** The material was based on two finite-state grammars, grammar A and B (Figure 1). Both were associated with the same tones (*a*, *a#*, *c*, *d*, *f#*) used to generate two sets of tone sequences, with none being shared between the grammars. Grammar A was based on Tillmann and Poulin-Charronnat (2010). Grammar B was created on the basis of grammar A, with an identical node structure, but reversed direction for two arrows and exchanged tone attributions, except for loops. See Figure 1 for example sequences of each grammar and for further explanations of sequence construction.

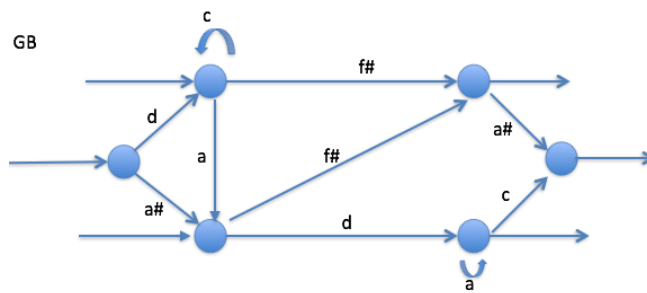
For each grammar, 24 grammatical sequences of four, five or six tones were created for the exposure phase. Each tone was played by an acoustic piano timbre and sounded for a duration of 500 milliseconds. The inter-tone interval was set to 0 ms. The resulting sequences were analyzed for their respective tone distributions (Table 1), interval distributions (Table 2) as well as their average pitch height. As can be seen in Table 1, the two grammars differed partly in their tone distributions, in particular with the tone *f#* being used more frequently in grammar A, while the tone *a* was used more frequently in grammar B. Regarding the interval distributions, Table 2 lists all possible intervals from tone repetition (coded as interval size of 0) up to 10 semitones, either ascending (+10) or descending (-10) for each grammar. The two distributions were compared by calculating the absolute differences between the two grammars' occurrences for each interval size, which were tested against 0 with a one-sample t-test. This test was significant ( $p = .00017$ ), suggesting different interval use between the two grammars. Regarding the average pitch of the exposure sequences, the five used tones



were ordered as a function of their pitch height. The lowest tone (*c*) was coded as 0 and the other tones were coded in reference to it, respecting the semitone differences (i.e., *d* → 2; *f*# → 6; *a* → 9; *a*# → 10). For each sequence, the tones were recoded with the appropriate pitch height code, allowing us to calculate an average pitch height for each sequence. The average pitch height for grammar A ( $4.26 \pm 2.49$ ) and grammar B ( $4.79 \pm 1.46$ ) did not differ significantly,  $p = .365$ . These analyses thus suggest that structural features allowing for distinguishing the two grammars were not related to pitch height differences, but rather mostly to the underlying interval distributions. Some of the intervals were more frequently used in one or the other grammar, such as the intervals +2, +3, -3 and -4 being used rather in Grammar A while the intervals -7, -9 and +9, being used rather in Grammar B. While this might suggest a use of slightly larger intervals in grammar B, this sole feature does not seem to be sufficient for distinguishing the two systems because grammar A also contained large intervals, grammar B also contained small intervals, and both grammars included tone repetitions. Consequently, the specificity of each grammar seemed to consist rather in the pattern of the overall tone distribution.

For the test phase, 24 additional grammatical sequences were constructed following the same implementation constraints, timing and material.





Example sequences:



Grammar A

Grammar B

Figure 1. Top) Presentation of the finite state grammar used for grammar A (top, adapted from Tillmann and Poulin-Charronnat, 2010) and grammar B (bottom). For each grammar, a set of nodes is connected to each other by arrows, and a sequence results from the chaining of the events between entering and exit points, following the arrows and their associated events. Bottom) One grammatical example sequence for each of the grammars: left: *d-f#-d-c* for grammar A; right: *d-a-d-c* for grammar B.

Table 1. Frequencies of occurrence of tones in exposure sequences of Grammar A and Grammar B

|           | Grammar   |           |
|-----------|-----------|-----------|
|           | Grammar A | Grammar B |
| <i>c</i>  | 36        | 36        |
| <i>d</i>  | 29        | 24        |
| <i>f#</i> | 23        | 9         |
| <i>a</i>  | 29        | 46        |
| <i>a#</i> | 8         | 8         |

Table 2. Frequencies of occurrence of pitch intervals in exposure sequences of Grammar A and Grammar B. (A) Pitch intervals were coded in semitones, all possible intervals were listed, starting with tone repetition (coded as interval size of 0) up to 10 semitones, either ascending (+10) or descending (-10). (B) For convenience, frequencies of occurrence were also presented independently of direction (i.e., ascending or descending), as referred to as absolute values (abs).

A)

|   | Grammar   |           |
|---|-----------|-----------|
|   | Grammar A | Grammar B |
| 0 | 22        | 28        |
| 1 | 2         | 0         |
| 2 | 14        | 0         |
| 3 | 8         | 0         |
| 4 | 5         | 5         |
| 5 | 0         | 0         |
| 6 | 0         | 4         |

|     |    |    |
|-----|----|----|
| 7   | 14 | 18 |
| 8   | 0  | 0  |
| 9   | 0  | 10 |
| 10  | 0  | 0  |
| -1  | 0  | 0  |
| -2  | 7  | 6  |
| -3  | 14 | 5  |
| -4  | 7  | 0  |
| -5  | 0  | 0  |
| -6  | 3  | 0  |
| -7  | 0  | 10 |
| -8  | 2  | 3  |
| -9  | 0  | 10 |
| -10 | 3  | 0  |

---

B)

|                | Grammar   |           |
|----------------|-----------|-----------|
|                | Grammar A | Grammar B |
| <i>abs(1)</i>  | 2         | 0         |
| <i>abs(2)</i>  | 21        | 6         |
| <i>abs(3)</i>  | 22        | 5         |
| <i>abs(4)</i>  | 12        | 5         |
| <i>abs(5)</i>  | 0         | 0         |
| <i>abs(6)</i>  | 3         | 4         |
| <i>abs(7)</i>  | 14        | 28        |
| <i>abs(8)</i>  | 2         | 3         |
| <i>abs(9)</i>  | 0         | 20        |
| <i>abs(10)</i> | 3         | 0         |

---

**Procedure.** Participants performed two exposure phases followed by test. In the exposure phases, the two grammars were presented in counterbalanced order across participants. For each exposure phase, participants performed a counting task on the exposure sequences, which were presented twice in random order. Participants were asked to count the number of times the first tone reappeared in the sequence (e.g., for *a#daf#a#*, the response is ‘two’). Exposure phases were separated by a short break when the experimenter informed participants that they continue the same task with new sequences.

After exposure, participants were told that the sequences they heard in the first phase were created by “composer 1” and those of the second phase by “composer 2”. They were told that each composer followed his own rules to create his sequences and that there was no sequence identical between the two composers. For each of the test sequences, participants were required to judge whether the sequence belonged to the repertoire of Composer 1 or Composer 2. After having given their response by clicking one of two virtual buttons on the screen, they judged their level of confidence using a subjective scale from 0 ‘I am guessing’ and 1 ‘I am not sure’ to 5 ‘I am totally sure’. Sequences were presented in random order to each participant. The stimuli were played at comfortable loudness level over headphones.

**Data analysis.** Test performance was coded as percentages of correct responses (average  $\pm$  standard deviation, Median (Md)). Tests for differences between grammars or orders and against chance were performed with two-tailed t-tests. To estimate chance level, we also ran Monte-Carlo simulation, which involved 2,100,000 replications answering randomly 48 items (Metropolis & Ulam, 1949). Simulations established chance level between 48% and 52% of correct responses. Exposure phase performance is reported with percentages of correct responses (average  $\pm$  standard deviation, Md), and is compared between grammar A and B with two-tailed t-tests. Normality was not violated in each of the conditions (Shapiro-Wilk tests). For significant effects, we also reported effect sizes as estimated with Cohen’s *d* (Cohen, 1988). We examined the zero correlation criterion following Dienes (2004, 2008) by calculating confidence ratings for correct and incorrect responses separately. For completion, we also performed Bayesian analysis to test performance against chance level. We used the software JASP (Wagenmakers et al., 2017), reporting Bayes Factor (BF<sub>10</sub>) as a relative measure of evidence. To interpret the strength of evidence, Lee and Wagenmakers (2014) suggested to consider a BF under three as weak evidence, a BF between three and 10 as

positive evidence, a BF between 10 and 100 as strong evidence and a BF higher than 100 as a decisive evidence.

## Results

**Test (Figure 2).** Participants reached 54.44 % ( $\pm 10.09$ ; Md=56.20) of correct responses, which was significantly different from chance ( $t(29) = 2.41, p = .02, d = .44$ ). Bayesian analyses revealed a  $BF_{10} = 2.30$ , suggesting weak, though relatively close to positive evidence for above chance performance. Mean performance did not differ between grammar A and B;  $t(29) = .42, p = .68$ , and according to presentation order,  $p = .82$ ; participants reached 54.01% ( $\pm 10.18$ ; Md=56.25) when the first exposure phase was grammar A and 54.88% ( $\pm 10.33$ ; Md=56.15), when it was grammar B. Performance did not differ between the first presented grammar and the second presented grammar (whether A or B),  $p = .41$ . Note that performance for grammar A correlated significantly with performance for grammar B ( $r(28) = .387, p < .05$ ), suggesting that participants who performed well in one grammar performed also rather well in the other grammar. According to Monte-Carlo simulation, all accuracy levels were above chance. Confidence ratings for correct responses ( $2.08 \pm 0.57$ ) did not differ from confidence ratings for incorrect responses ( $2.03 \pm 0.55$ ),  $p = .31$ .

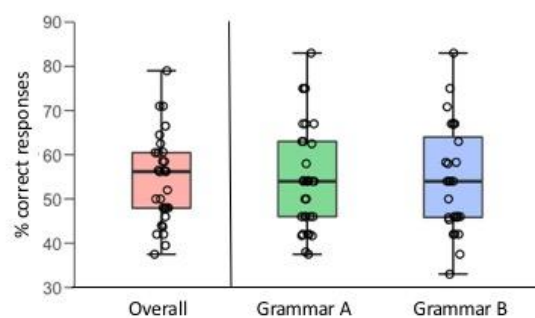


Figure 2. Box-plot representation of correct responses (in %) observed in Experiment 1 for overall performance as well as separated for Grammar A and Grammar B. The box indicates the interquartile range (with 75% and 25% quartiles as borders) and contains the median value, represented as a bar. In addition, circles represent individual data points of each participant, and error bars represent the range between maximum and minimum values

**Exposure.** Average accuracy (over both exposure phases) was 54.93% ( $\pm 16.86$ ; Md=51.56), and performance did not differ between grammar A and B ( $55.14 \pm 17.96$ , Md = 52.08, and  $54.72 \pm 19.44$ , Md = 52.08, respectively;  $p = .89$ ). Performance in exposure and test phases did not correlate ( $r(28) = 0.21$ ). These results suggest that participants paid attention to the exposure sequences, and similarly to both grammars.

## Discussion

Experiment 1 investigated whether participants can learn two AGs implemented with the same events in the same modality over two exposure phases. The test phase required participants to attribute new (grammatical) sequences to either composer 1 (first exposure phase, AG1) or composer 2 (second exposure phase, AG2). The above-chance-level accuracy in this categorization task shows that participants differentiated sequences of the two AGs. This finding suggests intra-modal learning for stimuli that share the same perceptual properties, here non-verbal auditory material. Analyses of the relation between accuracy and confidence ratings suggest that the acquired knowledge remained at an implicit level. The analyses of the grammatical sequences (see Method section) suggest that participants became sensitive to differences in interval distributions used in the two grammars. This observation is in agreement with findings of an implicit learning study investigating the acquisition of 12-tone musical materials (i.e., atonal music). After having been exposed to short musical

excerpts that respected a given “series” as defined in 12-tone music systems, participants were able to detect this series in a set of new excerpts. The material analyses revealed that participants became sensitive to the frequency distribution of tone intervals in the atonal musical material (Bigand et al., 2008; Bigand et al., in prep.). Our findings integrate into previous reports showing that, with exposure, listeners become sensitive to new structural systems based on tones (e.g., Loui et al., 2010; Rohrmeier et al., 2011) as well as more specifically to the distribution of tones (e.g., Oram & Cuddy, 1995) and tone transitions (e.g., more frequent specific intervals in Krumhansl et al. 1999). While the acquisition of tone distributions has been studied also with material that controlled for interval differences (Cui et al., 2016), less attention has been given up to now to the learning of interval distributions or tone chunk/transition learning independently of implemented pitch height. Interestingly, adults have been shown to be more sensitive to relative pitch patterns (i.e., tone patterns (intervals) described by the tones independently of pitch height) rather than absolute pitch patterns (which tones are played), while infants favor absolute pitch information (e.g., Saffran, 2003). Future research could now further investigate more systematically the learning of interval distributions, that is going beyond interval patterns (as in Saffran & Griepentrop, 2001) and controlling tone distributions, for the learning of one system or two systems.

Regarding our present results, it might be argued that participants were able to answer correctly even though they learned only one of the grammars, that is recognizing the style of one composer while rejecting anything else (i.e., without an understanding of the style of the other composer). Experiment 2 tested each of the AGs separately in two test phases. In each test, participants had to distinguish new grammatical sequences from sequences containing a grammatical violation. If participants are able to learn both AGs, test performance should be above chance level for each AG.

## Experiment 2

### Method

**Participants.** Thirty-one other participants took part in Experiment 2. 15 participants with a mean age of 22.93 years ( $\pm 4.56$ ) were first exposed to grammar A, then grammar B and 16 participants with a mean age of 24.31 ( $\pm 7.66$ ) years were first exposed to grammar B, then A. All participants indicated 0 years of musical training.

**Material.** For each grammar, the 24 exposure sequences of Experiment 1, 24 new grammatical test sequences and 24 ungrammatical sequences were used. For ungrammatical sequences, one tone of each grammatical sequence was replaced to create a tone succession that was not allowed by the grammar. Note that all replacements happened within the system (i.e., no new tones were introduced from outside the tone set used to create the initial items). For instance, the grammatical sequence *a#ccdc* was modified to *a#a#cdc* to become ungrammatical. We used this method as introduced in our earlier tone AG learning study (Tillmann & Poulin-Charronnat, 2010) with the aim to avoid the creation of new, previously unseen bigrams (i.e., bigrams that had not occurred in the exposure sequences) that preceded or followed the changed tones (a bigram is defined as a tone pair, such as *c-d*). However, after the study has been run, we discovered some construction errors that had led to the introduction of some new bigrams in the ungrammatical items, even though most of the items respected the initial constraints: New bigrams occurred for seven test items for Grammar A (out of 24 items) and four test items for Grammar B (out of 24 items). As it was the case in Tillmann and Poulin-Charronnat (2010), the creation of ungrammatical test items introduced new trigrams of tones (defined as a set of three successive tones, such as *d-a-a*) in almost all items, except for one test item in Grammar A and seven test items for Grammar B.

**Procedure.** Exposure was as described in Experiment 1. Then, the experimenter told participants that two test phases, one for each composer, follow. In the first test phase,



participants were told that they will hear new regular sequences of composer 1 and other sequences of composer 1 that contained an error. Participants classified each sequence as having been created by composer 1 or as having been modified. Participants also judged their level of confidence as in Experiment 1. For the second test phase, the same instructions were given, but related to composer 2. In each test phase, participants worked on 24 grammatical and 24 ungrammatical sequences, presented in random order.

The presentation order of the grammars for exposure determined the presentation order at test: For example, participants who heard grammar A followed by grammar B for exposure were first tested for grammar A and second for grammar B.

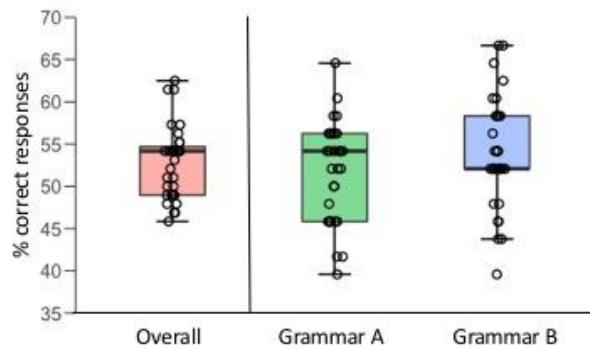
**Data analysis.** Exposure and test phase performance were analyzed as in Experiment 1, except that test performance was assessed against chance with one-sided t-tests as performance should be superior to 50% to reflect learning. Normality was not violated in each of the conditions (Shapiro-Wilk test). As accuracy in the test phase analysis averaged performance over grammatical and ungrammatical items, we complemented the analyses with the calculation of the response-bias-corrected indicator of [%Hits - %FalseAlarms]. Hits were defined as correct responses for ungrammatical items (correctly detecting the ungrammaticality) and False Alarms were defined as errors for grammatical items. For this indicator, chance performance would be 0. In addition, we used signal detection theory to calculate discrimination sensitivity with  $d'$  and response bias with  $c$  for each participant.  $d'$  is defined as  $z(\text{Hits}) - z(\text{FAs})$ , and response bias  $c$  as  $-0.5(z(\text{Hits}) * z(\text{FAs}))$ ; (see Macmillan and Creelman, 1991).

## Results

**Test (Figure 3).** Participants reached 52.86% ( $\pm 4.36$ ;  $Md=54.17$ ) of correct responses, which was significantly above chance ( $t(30) = 3.65, p < .0001, d = .655$ ). Bayesian analyses revealed a  $BF_{10} = 65.66$ , suggesting strong evidence for above chance performance. Mean performance did not differ between the grammars ( $p = .21$ ), both being above chance ( $t(30) = 1.695; p = .05, d = .304$  and  $t(30) = 3.263; p < .001, d = .586$  respectively). Average performance did not differ according to presentation order of the grammars,  $p = .92$ ; participants reached 52.78% ( $\pm 4.09$ ;  $Md=52.08$ ) when the first exposure phase was grammar A and 52.93% ( $\pm 4.74$ ;  $Md=54.17$ ), when it was grammar B. Performance did not differ between the first presented grammar and the second presented grammar (whether A or B),  $p = .91$ . According to the Monte-Carlo simulation (see Experiment 1), average overall performance was above chance level (also for performance of grammar B and at cut-off (52%) for grammar A).

Participants' performance as measured by the indicator of [Hits – False Alarms] was also significantly above chance level ( $6.11\% \pm 8.58$ ;  $t(30) = 3.901; p < .001, d = .712$ ), with  $47.92\% \pm 7.67$  of Hits and  $41.81\% \pm 7.03$  of False Alarms, when considered separately. SDT analyses revealed weak sensitivity, though significantly above 0 ( $p < .001; d' = 0.158 \pm .221$ ), and a positive response bias ( $c = .131 \pm .156$ ), which was significantly above 0 ( $p < .001$ ). The positive response bias reflects participants' tendency to respond grammatical, that is missing the ungrammaticality, which reflects the difficulty of the task.

Due to a programming error, confidence ratings were recorded only for 14 participants: confidence ratings for correct responses ( $2.113 \pm 0.620$ ) did not differ from confidence ratings for incorrect responses ( $2.107 \pm 0.626$ ),  $p = .91$ ; this was also the case for Grammar A and Grammar B considered separately,  $ps > .53$ .



**Figure 3.** Box-plot representation of correct responses (in %) observed in Experiment 2 for overall performance as well as separated for Grammar A and Grammar B. The box indicates the interquartile range (with 75% and 25% quartiles as borders) and contains the median value, represented as a bar. In addition, circles represent individual data points of each participant and error bars represent the range between maximum and minimum values

**Exposure phases.** Participants reached similar performance level as in Experiment 1 (54.47 %  $\pm$  12.18; Md=55.21), suggesting that they paid attention to the exposure sequences. In contrast to Experiment 1, performance was significantly better for grammar B than A,  $t(30) = 2.33$ ,  $p = .03$ . Performance in exposure and test phases did not correlate,  $r(29) = 0.09$ .

## Discussion

Experiment 2 aimed to further investigate the potential knowledge acquired for two AGs. Participants were required to judge whether the new sequences contained an error (regarding the composer's style) or not. Average accuracy in the test phase was above chance level and did not differ between the two AGs. Above chance-level performance was not only observed for overall accuracy (averaging across grammatical and ungrammatical items) but also for the

indicator of Hits-FalseAlarms, which correct for potential response biases of participants favoring one or the other response option. These results suggest that participants learned about regularities of each AG. In agreement with Experiment 1, analyses of the relation between accuracy and confidence ratings suggests that the acquired knowledge remained implicit, but we have to acknowledge that we were missing data for a subset of the participants in Experiment 2.

The relatively low test performance might not only be due to the learning of two systems, but also the test implementation. Test items were presented one by one, aiming to keep the presentation as in Experiment 1. Presenting test items by pair (pairing a new grammatical sequence with its ungrammatical one) in a 2AFC paradigm (as often used in AL research) might lead to increased performance.

### **General discussion**

Our findings show incidental learning of two tone AGs, presented in separate exposure phases. Participants categorized new items as belonging to one or the other grammar (Experiment 1) or detected grammatical errors in new sequences (Experiment 2). The task of Experiment 1 required the classification of new sequences as belonging to composer 1 or composer 2; a task similar to tasks used by Conway and Christiansen (2006) and Franco et al. (2011). The task of Experiment 2 allowing for finer knowledge testing suggest that participants did not succeed in Experiment 1 by only recognizing the style of one composer while rejecting anything else (i.e., without an understanding of the style of the other composer), but they become sensitive to the regularities of both composers.

The findings suggest that learning two artificial systems is (1) not restricted to AL, but extends to AG, and (2) does not require an additional surface cue (e.g., different voices, as implemented in Franco et al., 2011). Presenting the AGs in two exposure phases avoided potential confusion between the AGs based on the same tones. Also in agreement with Gebhart et al. (2009), these results call for future investigation of intermixed exposure phases and the potential influence of additional cues. Weiss et al. (2009) have shown that two ALs can be learned in an intermixed presentation with and without indexical voice cues. In their implementation, the two ALs alternated in substrings of two minutes, thus allowing for the presentation of a set of words of one AL before switching to the other AL. In contrast, Conway and Christianson (2006) changed AG with each string, thus leading to a quick alternation from one AG to the other. This difference leads to the hypothesis that presenting several items of one AG before presenting the items of the other AG might allow learning both AGs (without indexical cue or phase separation), as observed for ALs.

Despite above chance level performance, it might be argued that participants' performance in both tasks was rather low, leading to the question whether there might be some cost of learning two AG systems rather than learning only one AG system. Performance level in Experiment 2<sup>2</sup> can be compared with performance reported in a recent study from our lab (Poulin-Charronnat et al., in prep.): For a highly similar tone AG tested with a grammaticality judgment task, participants performed on average with an accuracy of 57.94 % (SD±7.15), which was also above chance level and also showing relatively high variability as observed here. Future research should now directly compare the learning of one versus two AG

---

<sup>2</sup> Note that performance level cannot be compared directly to performance level of Franco et al. (2011) or to Tillmann and Poulin-Charronnat (2010). Franco et al. used the AL paradigm where participants have to detect the words they have heard in the exposure phase (vs. non- or pseudo-words) while the AG paradigm (as used in our present study) requests participants to judge new (previously unheard) sequences as belonging to the same grammar or note. Tillmann and Poulin-Charronnat (2010) used an AG of tones that participants were exposed to, but learning was tested with an implicit task measuring response times for an out-of-tune detection on an either grammatical or ungrammatical target tone (and not grammaticality judgements, like here). For this implicit task, it was necessary to aim for high accuracy (participants' capacity to correctly judge a tone to be in-tune or out-of-tune) as the principal dependent variable was correct response times.

systems, aiming to determine the potential cost of learning two systems as well as to compare the inter-individual differences for the two types of learning situations.

An additional control condition could be testing participants without a preceding exposure phase. Albeit interesting, no-exposure controls have also been considered as problematic (see Perruchet & Reber, 2003), in particular because of potential within-test-phase learning (see Rohrmeier et al., 2011). Nevertheless, we had also implemented no-exposure control groups in two previous studies that used very similar tone AG exposure associated in the test phase with either classical grammaticality judgements (Poulin-Charronnat et al., in prep) or a priming task (Tillmann & Poulin-Charronnat, 2010). Both test phases without exposure confirmed no systematic bias in the material, with either at chance performance (Poulin-Charronnat et al., in prep) or no difference in response times (Tillmann & Poulin-Charronnat, 2010). However, the previously raised criticism regarding no-exposure control groups whose aim is to test for potential material biases emerging from the grammar implementation and its test items, points out the need to increase the control of both learning and test materials in implicit learning experiments. In studies investigating the learning of a single AG, learning and test materials have been often constructed with some a priori constraints combined with some a posteriori analyses of the material, as we did here (e.g., Udden & Männel, 2018; Poletiek, Van Schijndel, 2009). As the control of the materials becomes even more important when testing the learning of two systems, future studies need to improve the artificial systems implementation. For this goal, computational modelling and statistical analyses approaches can help generating and selecting stimulus materials, analyzing potentially correlated features and avoiding potential confounds as well as analyzing and interpreting the results (e.g., Zuidema et al., 2019; Bailey & Pothos, 2008; Pothos, 2010).

Our study aimed for an implementation favoring implicit processing, which has been shown to allow for more complex learning than explicit processing (e.g., Howard & Howard,

2001; Fletcher et al., 2005). First, we used the cover story of two composers, allowing us to avoid telling participants about underlying grammars at exposure and test. Second, we implemented the AG not with verbal material, but with tone material that is less subjected to explicit strategies. The results of the zero-correlation criterion (Dienes, 2004, 2008) provide some first indication that participants' knowledge might remain at an implicit level. Future research needs now to further investigate in how far the knowledge remains indeed implicit (e.g., Dienes & Perner, 1999; Shanks & St John, 1994; Seger, 1994), or reaches some level of explicitness, notably by using (1) additional tests, which have been previously proposed in implicit learning research (e.g., combined with the Serial Reaction Time task in particular) to investigate the implicitness versus explicitness of the acquired knowledge (e.g., questionnaires and familiarity ratings (e.g., Terry et al., 2016) as well as generation tasks with exclusion/inclusion criteria (e.g., using an adapted process dissociation procedure <sup>3</sup>, Destrebecqz & Cleeremans, 2001; Schultz et al., 2013), and (2) more implicit tasks also in the test phase, such as the priming task, which provides an indirect investigation method without revealing explicit material features to the participants throughout the experiment (Vinter & Perruchet, 1999; Tillmann & Poulin-Charronnat, 2010).

In sum, our findings suggest the capacity of intra-modal learning of regularities in the auditory modality and for stimuli sharing the same perceptual properties (two AGs based on the same tone set), opening up to perspectives for further investigating the cognitive capacity of implicit learning, which might indeed be more powerful than sometimes observed.

---

<sup>3</sup> After the exposure phase, participants are required to generate patterns under two types of instruction: (a) an inclusion instruction asking to reproduce the exposure pattern, and (b) an exclusion instruction asking to create new patterned sequences. The sequences produced under both instructions are compared to the exposure pattern. If similarity in the inclusion instruction is inferior to or equal to similarity in the exclusion instruction, results suggest implicit knowledge. However, higher similarity scores in the inclusion instruction than the exclusion instruction suggest explicit knowledge.

## **Acknowledgements**

We thank G. Jeantet, *Meludia*, the ministerial program of ASRT as well as Alexandra Corneyllie for support in the present project. This work was conducted in the framework of the LabEx CeLyA (“Centre Lyonnais d’Acoustique”, ANR-10-LABX-0060) of the University of Lyon within the program “Investissements d’avenir” operated by the French National Research Agency (ANR).

## **Declaration of conflicting interests**

The authors declare that there is no conflict of interest.



## References

- Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 899–912.
- Bailey, T.M., Pothos, E.M. (2008). AGL StimSelect: Software for automated selection of stimuli for artificial grammar learning. *Behav Res* 40, 164–176.
- Bigand, E., D'Adamo, D. & Poulin, B. (in prep). The implicit learning of 12-tone music. *Manuscript under preparation*.
- Bigand, E., Lalitte, P. & Tillmann, B. (2008). Learning music: prospects about implicit knowledge in music, new technologies and music education. (pp. 47-81) In: Polotti, P. & Rocchesso, D. (Eds). *Sound to Sense: Sense to Sound*, Logos Verlag: Berlin.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities pitting abstract against stimulus-specific representations. *Psychological Science*, 17(10), 905-912
- Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin and Review*, 8(2), 343–350.
- Dienes, Z., Altmann, G. T. M., Kwan, L. & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 21, 1322-1338.
- Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, 22, 735–808.

Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies*, 11, 25-45.

Dienes, Z. (2008). Subjective measures of unconscious knowledge. *Progress in Brain Research*, 168, 49-64.

Emberson, L. L., Conway, C. M. & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning, *The Quarterly Journal of Experimental Psychology*, 64, 1021-1040.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504.

Fiser, J. & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 28, 458–467.

Fletcher, P.C., Zafiris, O., Frith, C.D., Honey, R.A.E., Corlett, P.R., Zilles, K. & Fink, G.R. (2005). On the Benefits of not Trying: Brain Activity and Connectivity Reflecting the Interactions of Explicit and Implicit Sequence Learning. *Cerebral Cortex*, 15, 1002-1015.

Franco, A., Cleeremans, A. and Destrebecqz, A. (2011). Statistical learning of two artificial languages presented successively: how conscious? *Front. Psychology*, 2, 229.

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing Structures in Midstream: Learning Along the Statistical Garden Path, *Cognitive Science*, 33, 1087-1116.

Hoch, L., Tyler, M.D., & Tillmann, B. (2013). Regularity of unit length boosts statistical learning in verbal and nonverbal artificial languages. *Psychonomic Bulletin and Review*, 20(1):142-7.

Howard, D.V., & Howard, J.H., Jr (2001). When it does hurt to try: adult age differences in the effects of instructions on implicit pattern learning. *Psychonom Bull Rev.*, 8, 798–805.

Hunt, R. H. & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, 130, 658-680.

Krumhansl, C., Toivanen P., Eerola, T., Toiviainen, P., Järvinen, T., and Louhivuori, J. (2000) Cross-cultural music cognition : cognitive methodology applied to North Sami yoiks, *Cognition* 76, 13-58.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge Core - Psychology Research Methods and Statistics.

Loui, P., Wessel, D., Kam, CLH (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music perception* 27 (5), 377-388.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

McMullen, E. & Saffran, J. R. (2004). Music and language A developmental comparison. *Music Perception*, 21, 289-311.

Opacic T, Stevens K, Tillmann B. (2009) Unspoken Knowledge: Implicit Learning of Structured Human Dance Movement. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 35, 1570-7.

Oram, N., & Cuddy, L.L. (1995). Responsiveness of Western adults to pitch distributional information in melodic sequences. *Psychological Research* 57:103–118.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10(5), 233–238.

Perruchet, P. (2008). Implicit learning. In H. Roediger (ed.), *Cognitive psychology of memory*. Vol. 2 of learning and memory: A comprehensive reference (pp. 597–621). Oxford, England: Elsevier.

Poletiek, F. H. & Van Schijndel, T. J. P. (2009). Stimulus set size and statistical coverage of the grammar in artificial grammar learning. *Psychonomic Bulletin & Review*, 16 (6), 1058-1064

Pothos EM (2010) An entropy model for artificial grammar learning. *Front. Psychology* 1:16. doi: 10.3389/fpsyg.2010.00016.

Prince, J., Stevens, K., Jones, M. & Tillmann, B. (2018). Implicit learning of an artificial grammar structure in pitch and/or time. *Journal of Experimental Psychology: Learning Memory and Cognition*, 44(8):1201-1214.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6, 855-863.

Rebuschat, P. & Rohrmeier, M. (2012). Implicit learning and acquisition in music. *Topics in Cognitive Science*, 4, 525–553.

Rohrmeier, M., Rebuschat, P. & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness and Cognition*, 20, 214–222

Saffran, J.R. (2001). The use of predictive dependencies in language learning. *Journal of Memory & Language*, 44, 493-515.

Saffran, J. R. (2003). Absolute pitch in infancy and adulthood: the role of tonal structure. *Developmental Science*, 6:1, 35– 47.

Saffran, J.R., & Griepentrog, G.J. (2001). Absolute pitch in infant auditory learning: evidence for developmental reorganization. *Developmental Psychology*, 37, 74 – 85.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 8, 101–105.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27-52.

Seeger, C. A. (1994). Implicit learning. *Psychological Bulletin*, 115, 163-196.

Schultz, B., Stevens, K., Keller, P. & Tillmann, B. (2013). The Implicit Learning of Metrical and Non-metrical Temporal Patterns. *Quarterly Journal of Experimental Psychology*, 66(2):360-80.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17, 367–447.

Vinter, A., & Perruchet, P. (1999). Isolating unconscious influences: The neutral parameter procedure. *The Quarterly Journal of Experimental Psychology*, 52A, 857–875.

Terry, J., Stevens, K., Weidemann, G., & Tillmann, B. (2016). Implicit learning of between-group intervals in Auditory Temporal Structures. *Attention, Perception & Psychophysics*, 78(6):1728-43.

Tillmann, B. & McAdams, S. (2004) Implicit Learning of musical timbre sequences : statistical regularities confronted with acoustical (dis)similarities. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 30(5),1131-1142.

Tillmann, B. & Poulin-Charronnat, B. (2010). Auditory expectations for newly acquired structures. *The Quarterly Journal of Experimental Psychology*, 63, 1646-1664.

Udden J, & Männel C (2018) Artificial grammar learning and its neurobiology in relation to language processing and development. In: Oxford: Oxford University Press., editor. *Oxford handbook of psycholinguistics*. pp. 755–783.

Vinter, A., & Perruchet, P. (1999). Isolating unconscious influences: The neutral parameter procedure. *The Quarterly Journal of Experimental Psychology*, 52A, 857–875.

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 1–19.

Weiss, D. J., Gerfen, C. & Mitchel, A. (2009). Speech segmentation in a simulated bilingual environment: challenge for statistical learning? *Lang. Learn. Dev.*, 5, 30-49.

Zuidema, W., French, R. M., ALhama, R. G., Ellis, K., O'Donnell, T. J., Sainburg, T. & Gentner, T. Q. (2019). Five Ways in Which Computational Modeling Can Help Advance Cognitive Science: Lessons From Artificial Grammar Learning. *Topics in Cognitive Science*, 1–17.