NeuroImage xxx (xxxx) xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

From learning to creativity: Identifying the behavioural and neural correlates of learning to predict human judgements of musical creativity

Ioanna Zioga [a,*], Peter M.C. Harrison [b], Marcus T. Pearce [b,c], Joydeep Bhattacharya [d], Caroline Di Bernardi Luft [a,**]

[a] Department of Biological and Experimental Psychology, School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NS, United Kingdom
[b] School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS, United Kingdom
[c] Centre for Music in the Brain, Aarhus University, Aarhus, Denmark
[d] Department of Psychology, Goldsmiths, University of London, London, SE14 6NW, United Kingdom

ARTICLE INFO

Keywords:
Creativity
Artificial music grammar
EEG
Statistical learning
Training
IDyOM

ABSTRACT

Human creativity is intricately linked to acquired knowledge. However, to date learning a new musical style and subsequent musical creativity have largely been studied in isolation. We introduced a novel experimental paradigm combining behavioural, electrophysiological, and computational methods, to examine the neural correlates of unfamiliar music learning, and to investigate how neural and computational measures can predict human creativity. We investigated music learning by training non-musicians ($N = 40$) on an artificial music grammar. Participants' knowledge of the grammar was tested before and after three training sessions on separate days by assessing explicit recognition of the notes of the grammar, while additionally recording their EEG. After each training session, participants created their own musical compositions, which were later evaluated by human experts. A computational model of auditory expectation was used to quantify the statistical properties of both the grammar and the compositions. Results showed that participants successfully learned the new grammar. This was also reflected in the N100, P200, and P3a components, which were higher in response to incorrect than correct notes. The delta band (2.5–4.5 Hz) power in response to grammatical notes during first exposure to the grammar positively correlated with learning, suggesting a potential neural mechanism of encoding. On the other hand, better learning was associated with lower alpha and higher beta band power after training, potentially reflecting neural mechanisms of retrieval. Importantly, learning was a significant predictor of creativity, as judged by experts. There was also an inverted U-shaped relationship between percentage of correct intervals and creativity, as compositions with an intermediate proportion of correct intervals were associated with the highest creativity. Finally, the P200 in response to incorrect notes was predictive of creativity, suggesting a link between the neural correlates of learning, and creativity. Overall, our findings shed light on the neural mechanisms of learning an unfamiliar music grammar, and offer novel contributions to the associations between learning measures and creative compositions based on learned materials.

1. Introduction

Human creativity is linked to acquired knowledge. Learning has been widely considered a statistical process during which humans learn, through exposure, the regularities of hierarchical structures, such as music and language (Rohrmeier and Koelsch, 2012). The overarching question of our study is about the relationship between statistical learning and musical creativity. What are the neural signatures of learning an unfamiliar music grammar? What differentiates an individual who learns well from another who does not? What are the characteristics of musical compositions that are judged to be highly creative? In the current study, we investigated these questions. Non-musicians were trained on an unfamiliar musical grammar and subsequently produced their own musical compositions. We adopted a tripartite approach by combining behavioural, neural, and computational methods to assess associations between statistical learning and creativity.

Humans operate as probabilistic inference machines that are continually trying to build internal probabilistic models of the outside world,

* Corresponding author.
** Corresponding author.
E-mail addresses: i.zioga@qmul.ac.uk (I. Zioga), c.luft@qmul.ac.uk (C. Di Bernardi Luft).

https://doi.org/10.1016/j.neuroimage.2019.116311
Received 31 May 2019; Received in revised form 18 October 2019; Accepted 22 October 2019
Available online xxxx

which they use to interpret incoming sensory information (see Rohrmeier and Koelsch, 2012, for a critical review on musical prediction and probabilistic learning). Statistical learning refers to the psychological mechanisms by which individuals learn the statistical properties of a particular sensory domain (e.g., language, music, visual sequences) through exposure (see Christiansen, 2019, for a definition and historical survey of cognitive scientific research on statistical learning). Usually statistical learning takes place through acquisition of knowledge according to transitional probabilities between the elements of rule-based structures or patterns (Saffran et al., 1999). Through this cognitive mechanism we extract the underlying patterns even after only a short exposure to stimuli (Lieberman et al., 2004; Loui, 2012; Luft et al., 2016; Misyak et al., 2010; Pothos, 2007; Reber, 1993; Rohrmeier and Cross, 2014; Rohrmeier and Rebuschat, 2012; Saffran et al., 1996). As probabilistic relationships between elements can easily be quantified and measured computationally, statistical learning offers a well-controlled experimental paradigm. In the auditory domain, statistical learning has been demonstrated in tone sequences (Saffran et al., 2005) and timbre sequences (Tillmann and McAdams, 2004). In order to eliminate potential effects of familiarity with the Western musical culture, unfamiliar, non-Western scales have been also used (e.g., Loui and Wessel, 2008; Loui et al., 2010). For example, Loui et al. (2010) used short melodies generated by an artificial and non-octave-based musical system, the Bohlen-Pierce scale, and found evidence for acquired knowledge and increased preference for this system as early as after 25 min of exposure.

Both behavioural and neurophysiological measures have been used to identify the neural signatures of statistical learning. The N100 component has been reported in response to tones with low transitional probability (i.e., unpredictable tones) compared to those with high transitional probability (i.e., predictable tones) (Abla et al., 2008; Abla and Okanoya, 2009; Carrión and Bly, 2008; Mandikal Vasuki, Sharma, Ibrahim and Arciuli, 2017). Its magnitude has also been positively correlated with the degree of expectancy: it is inversely modulated by the probability of the notes, i.e. the lower the probability, the higher the N100 magnitude (Koelsch et al., 2016). Of note, the N100 component is usually associated with violation of melodic expectations in Western tonal sequences (Koelsch and Jentschke, 2010; Pearce et al., 2010b). Another early component, the P200, has been linked to perception of deviant, invalid stimuli (e.g., vision: Gruber and Müller, 2004; audition: Freunberger et al., 2007), representing a mismatch between the sensory stimulus and the expected event. Violation of expectations has been also linked to later components, such as the P3, a positive ERP peaking around 300–600 ms after the onset of an unpredictable event (Arthur and Starr, 1984; Knight et al., 1989; Yamaguchi and Knight, 1991), interpreted as a detection of novelty (Polich, 2007).

A promising way to investigate the relationship between learning and creativity would be to consider not only the statistical nature of learning, but also the statistical features of the creative products. Creativity is usually defined as the generation of novel, surprising and valuable ideas (Runco and Jaeger, 2012). One way to generate creativity is through "the generation of unfamiliar combinations of familiar ideas" (Boden, 2010), and these unusual combinations are often interesting, thought-provoking, and humorous, but also require deep knowledge and flexible cognitive skills in order to form links between different concepts. In addition to such combinatorial creativity, Boden (2010) proposed that exploration of new parts of a conceptual space is likely to produce creative artefacts, as they are surprising, i.e. they cannot be predicted in advance (Simonton, 2012). Surprise is critically relevant in the context of music, as unusual transitions between elements (e.g., notes/chords) are unexpected and thus surprising to the listener (Narmour, 1992). Therefore, violation of expectations constitutes a crucial component of the emotional and aesthetic experience of music (Huron, 2006; Juslin, 2019). In his work on modelling surprise in jazz harmonic progressions, Pachet (1999) has further underlined the importance of building up expectations for interesting and exciting surprises to occur.

Following the standard definition of creativity, Boden's (2010, 2004) concept of idea generation, and framing our approach in a statistical context, an artefact is considered creative if it has low probability of occurrence (novel) and if it is correct according to a given grammar (adequate). We therefore hypothesized that creativity, in a statistical learning context, arises from the use of grammatically correct but unusual combinations of elements; incorrect elements would dissatisfy the requirements of creativity, as they do not belong to a given structure, while highly probable elements would be too obvious and conventional, therefore not creative either.

Neuroimaging studies on musical creativity have investigated the effect of musical expertise on the neural correlates of improvisation (Bengtsson et al., 2005; Limb and Braun, 2008; Liu et al., 2012; Lopata et al., 2017; for a review see Beaty, 2015). In an fMRI study, Limb and Braun (2008) asked professional jazz pianists to either perform a memorized melody or freely improvise over a pre-recorded rhythm. Improvisation was correlated with enhanced activity in medial prefrontal regions and reduced activity in lateral prefrontal regions. Similar results were found in a vocal analogue of this study with professional freestyle rap artists (Liu et al., 2012). In line with these findings, Pinho et al. (2014) found that during improvisation, experienced improvisers showed reduced activity in the right dorsolateral prefrontal cortex and inferior frontal gyrus. Overall, these results suggest less involvement of inhibitory processes and enhancement of stimulus-independent cognitive processes during musical improvisation of experts.

To date, creativity is assessed mostly by human experts. The most widely used method of assessing creative products is the Consensual Assessment Technique, CAT (Amabile, 1982). CAT makes use of the idea that the best way of assessing a creative product is by a committee of experts, trusting their own subjective experience to evaluate how creative the work is. The experts, working independently from each other and also blind to any experimental manipulation, rate each artefact on a creativity scale, and these ratings are averaged to produce an overall creative score (see Pearce and Wiggins, 2007, for an application of this technique to assess musical creativity). Interestingly, human experts are still required to evaluate creativity even when an artefact is generated by an algorithm (e.g., Varshney et al., 2013). However, little is known yet as to which statistical features of the products are characterized as highly creative by humans.

Therefore, the primary aim of this research was to draw a direct link between learning and creativity by first identifying the neural correlates of learning an unfamiliar musical grammar, and then investigating how the statistical features of musical compositions based on this grammar could predict creativity as judged by human experts. We tracked neural activity during first exposure to the musical material as well as during a retrieval phase. A computational model of auditory expectation, Information Dynamics of Music, IDyOM (Pearce, 2018) was used to compute the statistical properties of both the artificial music grammar, as well as those of participants' musical compositions. We thus attempted to find associations between learning and creativity on a statistical level.

Specifically, we introduced a novel experimental paradigm combining behavioural, electrophysiological, and computational methods to investigate the neural correlates of music learning and how they relate to subsequent generation of new music. In particular, we attempted to simulate music learning in real-life, by training non-musicians on an unfamiliar artificial music grammar, AMG (taken from Rohrmeier et al., 2011), over three training sessions on three separate days. This AMG is a finite-state grammar with tones belonging to the diatonic scale, constructed such that eight different tone pairs are combined under specific rules producing 18 different melodic sequences. An AMG was ideal for our purpose because it represents a completely novel musical style for all participants. Training included both passive exposure and an active reproduction of the melodies on a keyboard. Participants' knowledge of the AMG was assessed before and after training by testing the recognition of correct vs. incorrect notes. Their EEG was recorded during the test sessions (before and after training). After each training session, we asked participants to create and perform a musical

composition based on the newly acquired musical knowledge; these compositions were subsequently evaluated by human experts, as well as fed to IDyOM to extract information theoretic measures. In contrast to the previous studies investigating the effects of learning (Abla et al., 2008; Mandikal Vasuki et al., 2017), we attempted to identify the neural correlates of acquisition and retrieval of the newly learned material. Further, we explored how the signatures of learning (neural and behavioural) influence creativity, and, finally, how computational and EEG measures can predict human judgements of creativity.

Learning statistical properties of the auditory environment helps foster expectations for future events (Dienes and Longuet-Higgins, 2004; Loui et al., 2010). In our study, we used IDyOM to quantify the conditional probability of each note in every sequence, reflecting the expectedness of a given note given the preceding context. IDyOM uses variable-order Markov models (Begleiter et al., 2004; Bunton, 1997) to make predictions for forthcoming events from the frequency with which each event has followed the context in a reference corpus of music. IDyOM embodies the hypothesis that listeners base their expectations on learning statistical regularities in the musical environment, such that high-probability notes are perceived by listeners as expected, and low-probability notes as unexpected. Previous behavioural, physiological, and EEG studies have validated IDyOM's ability to predict listeners' expectations (Egermann et al., 2013; Pearce et al., 2010a).

Our hypotheses were as follows. At the behavioural level, we predicted that after training participants would be able to demonstrate learning of the novel grammar by recognizing note combinations which belong to the learned material. At the neural level, we first predicted that after training, as an index of learning, the N100, P200, and P3 amplitudes would be higher in response to incorrect compared to correct notes according to the AMG. With regards to the musical compositions that participants created, we hypothesized that (i) the more the correct notes used in a composition, the higher the perceived creativity, as incorrect notes would sound inappropriate to the judges, and (ii) note probability would have an inverted U-shape relationship with creativity, as medium probability notes would be considered as highly creative, while extremely high probability notes would be too expected and low probability notes too inappropriate. About the relationship between learning and creativity, we hypothesized that participants who learned the unfamiliar music grammar better would produce more creative musical compositions than the low learners. Furthermore, in this study, we explored the neural correlates of encoding vs. retrieval of the learned material, and contrasted how these two distinct mechanisms predict learning accuracy and creativity. We also investigated which quantitative features (neural mechanisms of learning, learning accuracy, computational features of compositions) made some musical compositions be judged by human experts as more creative than others. Finally, we explored what brain signatures of learning indicate whether a person produces creative musical compositions.

2. Methods

2.1. Participants

Forty neurologically healthy human adults (24 female) aged between 20 and 32 years old (mean ± s.d. age of 22.42 ± 3.04 years) participated in the experiment. Participants were non-musicians, as self-reported and validated by the 'Goldsmiths Musical Sophistication Index' (Gold-MSI) questionnaire (Müllensiefen et al., 2014). The Gold-MSI Musical Training questionnaire has a possible score of 7–49 points, and participants had a mean ± s.d. musical training score of 12.09 ± 4.60. Two participants were excluded from all analyses because they did not engage with the task (gave the same response throughout both test sessions), leaving 38 participants in total. All participants had self-reported normal hearing and normal or corrected-to-normal vision. Participants received monetary reimbursement at a rate of £7 per hour and gave written informed consent before the beginning of the experiment. The study protocol was approved by the Ethics Board at the Queen Mary University of London.

2.2. Materials

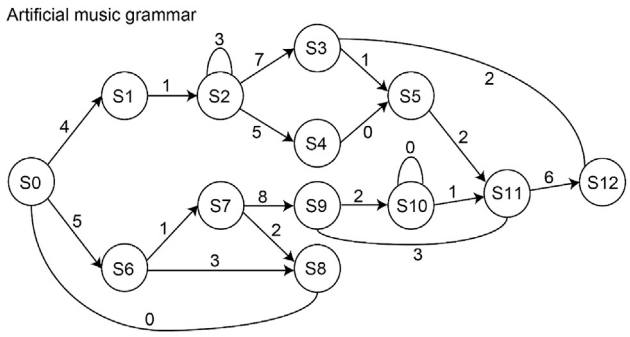
Gold-MSI Musical Training questionnaire: The Musical Training factor (Dimension 3) of the Gold-MSI was used to validate that participants did not have musical expertise. This self-report measure includes seven statements regarding formal musical training experience and musical skill. Each statement (e.g., '*I would not consider myself a musician*') requires a response on a 7-point scale ranging from 1 (*Completely Disagree*) to 7 (*Completely Agree*).

Artificial music grammar (AMG): We used melodic sequences generated by an artificial music grammar (AMG) proposed by Rohrmeier et al. (2011). This is a finite-state grammar with 8 different tone pairs as terminals (Fig. 1). The tones belong to the diatonic major scale, the most common scale in Western tonal music (C4, D4, E4, F4, G4, A4, and B4). The artificial music grammar generated 18 different melodic sequences, which were 8–22 notes long (mean length ± s.d.: 14.56 ± 3.87). Sequences with circular paths were excluded, as they were too long to be used in the test sessions, and also repetition was out of our area of investigation. Twelve of these sequences (old-grammatical) were used for the training and test sessions, while the remaining six (new-grammatical) were only presented in the last session of the experiment to test generalization to unseen examples from the grammar.

IDyOM analyses of melodic sequences: We configured IDyOM to analyse the statistical properties of the stimulus sequences generated by the AMG, as well as the statistical properties of the musical compositions produced by the participants. IDyOM's predictions are typically represented as *information content* (IC), defined as the negative log-probability of the next event, as estimated by the model on the basis of the preceding context. High (low) values of IC correspond to low (high) predictability. In our stimulus sequences, notes with IC in the lowest 30% of the distribution were categorised as high-probability while notes with IC in the highest 30% of the distribution were categorised as low-probability. Besides the high- and the low-probability notes, in the melodic sequences we used for the test sessions, we manipulated notes of the AMG to make them 'incorrect' (i.e. not permitted under the grammar). We also included 'random' sequences (i.e. all notes of a melody were randomly selected from the alphabet of possible notes). Participants' learning was evaluated in terms of their accuracy in recognizing that high-probability and low-probability notes belong to the AMG (i.e. are correct), whereas incorrect notes do not belong to the AMG (i.e. are incorrect). IDyOM was also used to estimate the probabilities of the notes of participants' musical compositions, in order to understand how they used their knowledge of the AMG to construct melodies.

Participants were presented with melodies ending with high-probability or low-probability notes with respect to the AMG, as calculated using the IDyOM model (Pearce, 2018, 2005), and asked to make judgements about these final notes.

Given a sequence x_1^k of length k, comprising elements $x \in X$, drawn from an alphabet X, IDyOM estimates the conditional probability of each element, given the preceding sequential context: $p(x_i|x_1^{i-1}), \forall i \in \{1 \ldots k\}$. Rather than using a fixed context length (or *order*) to estimate the probabilities, IDyOM computes a mixture of the probabilities estimated at different orders using a process known as smoothing (Chen and Goodman, 1999). The probabilities are estimated based on the frequency with which each element has been experienced in a given context during the training given to the model. IDyOM combines two models with different training regimes: the *short-term model* is trained incrementally from an initially empty state on the individual sequence currently being predicted while *the long-term model* has prior training on a larger corpus of sequences. The distributions generated by the long- and short-term models are combined using an entropy-weighted geometric mean (Pearce et al., 2005). IDyOM is also capable of combining distributions generated by models trained on different representations of the musical



Example sequence with IDyOM probabilities

Interval	4		1		7		1		2		6	
Note	F	A	E	F	B	E	E	F	A	G	A	E
Probability	2.53	1.88	0.86	0.16	3.28	0.89	3.30	0.30	0.77	0.77	1.06	0.20

Fig. 1. Top: An illustration of the artificial music grammar (AMG) developed and presented by Rohrmeier et al. (2011). Numbers 0–8 represent the musical intervals, and symbols starting with 'S' represent the nodes of the grammar. The nodes are points which connect the elements of the grammar (i.e. the musical intervals) with each other. Grammatical sequences start from the leftmost node and move along the pathways indicated by the arrows, until the rightmost node is reached. For example, one grammatical sequence starting from node S0 can move to S1, corresponding to interval 4 (F–A), and then to S2, corresponding to interval 1 (E–F), etc., until it reaches S12. Middle: Each interval corresponds to a pair of musical notes. Bottom: An example of a sequence used as melodic stimulus, including the intervals, the notes the intervals correspond to, and their probabilities, as estimated by a computational model of auditory expectation, IDyOM (Pearce, 2018). All musical notes are drawn from the range C4–B4.

surface (e.g., pitch, pitch interval, contour, scale degree). The distributions generated by these component models are also combined using an entropy-weighted geometric mean.

We analysed the information-theoretic properties of the 18 novel melodic sequences using IDyOM. The analyses used leave-one-out cross-validation, with IDyOM generating predictions for each sequence after pretraining on the other 17 sequences. IDyOM generates predictions using *viewpoints*, reflecting the various psychological features that listeners extract from melodies. We evaluated three candidate viewpoint sets that seemed plausible for the present study and selected the one that best predicted the melodic sequences by minimizing the cross-entropy error metric (Pearce, 2005). The best performing set comprised the viewpoints chromatic pitch and chromatic interval (cross-entropy = .986) and outperformed the single viewpoint chromatic pitch (cross-entropy = 1.007), and the viewpoint set chromatic pitch, chromatic interval, and contour (cross-entropy = 1.043).

By default, IDyOM makes predictions that combine a long-term model (representing stylistic knowledge) with a short-term model (representing statistical regularities learned from the current melody). The long-term model is pretrained on the 17 other melodies, and incrementally on the current melody; the short-term model is only trained incrementally on the current melody. This configuration (termed BOTH+) has been shown to reflect listeners' expectations well (Pearce, 2005), motivating its use in the present study. IDyOM returns a probability estimate for each note in each of the 18 melodies generated by the AMG. In line with previous work, we transformed this probability estimate by taking the negative logarithm (base 2), to produce *information content* (IC). High IC corresponds to low-probability, while low IC corresponds to high-probability notes, i.e. unexpected and expected notes, respectively, based on a specific grammar.

Additional analyses demonstrating that the AMG diverges from the Western musical style is included in Supplementary materials ("Comparison of the AMG with Western-pretrained model").

Melodic stimuli for judgement sessions: We assessed learning by asking participants to judge whether specific notes were correct or incorrect in the pre-test and post-test training sessions. In the last generalisation session, participants were asked to judge whether the last note was surprising or not surprising.

In the pre-test and post-test sessions, participants were presented with a total of 280 melodies, terminating with notes of different levels of information content (IC): 70 high-probability, 70 low-probability, 70 incorrect, and 70 random. For the generalisation session there were 105 melodies: 35 high-probability, 35 low-probability, and 35 incorrect. The melodies for the test sessions stemmed from the 12 old-grammatical sequences, whereas for the generalisation session they stemmed from the 6 new-grammatical sequences.

To generate the high-probability (HP) and low-probability (LP) melodic stimuli, we followed the following procedure: of all the notes of the 18 grammatical sequences, we identified those with IC in the lowest 30% of the distribution (extreme high-probability) and the notes in the highest 30% (extreme low-probability). These were the notes participants were asked to make a judgement about, and the presentation of the melody was interrupted after they were heard. There were 79 notes with extreme probabilities (high or low) in total, of which 55 belonged to the old-grammatical, and 24 to the new-grammatical sequences. Of the 55 old-grammatical, 36 notes were HP and 19 LP. To reach the 70 trials per condition, we repeated a number of melodies as appropriate. In particular, 34 (randomly picked) of the 36 HP melodies were repeated once, while all 19 LP were repeated three times (giving 57 melodies), and then 13 (randomly selected from the middle 40% of the distribution) were added (giving a total of 70). The same procedure was followed for the new-grammatical sequences. The 16 HP melodies were repeated once (32) and 3 (randomly selected from the middle 40% of the distribution) were added (giving 35 in total), while the LP 8 melodies were repeated four times (32) and 3 (randomly selected from the middle 40% of the distribution) were added (35).

To generate the incorrect melodies, we used the stems from the HP and LP melodies, but replaced the last note with a note that never appeared in the context of the AMG. We created three different sets of incorrect melodies, one for the pre (70), one for the post test (70), and one for the pleasantness judgements (35). Finally, we generated two different sets of 70 random melodies, for the pre and post sessions. Care was taken so that the random melodies had similar length to the rest of the melodies. To achieve this, we generated 5 random melodies for each of the possible lengths (7–20 notes).

The melodic stimuli were played from two speakers simultaneously. Each note lasted 330 ms and had a piano timbre. Psychtoolbox (Brainard and Vision, 1997), a MATLAB Toolbox was used to present the stimuli.

2.3. Procedure

The experiment took place on four separate days, with a maximum two-day gap between any two consecutive testing days (Fig. 2). On days 1–3, participants were trained on the melodies, created according to the AMG, through active reproduction on a keyboard. Participants' learning of the melodic sequences was tested before and after training (days 1 and 4): they were presented with melodies and asked to judge if the final note was correct or incorrect and surprising or not surprising, while their EEG signals were recorded. In the generalisation session, participants were asked to judge if the final note of new grammatical sequences was surprising or not surprising. Before training on days 2 and 3, participants attended a brief (5 min) passive exposure session where all old-grammatical sequences were presented three times (36 in total), and were then asked to complete a short surprisal (yes or no) judgement task of melodies ending with high-probability or low-probability notes

(intermediate surprisal sessions). After each training session, participants were asked to compose and perform a musical composition.

After the end of the data collection, four professional musicians were recruited as judges to evaluate the musical compositions. First, they received training on the AMG through passive listening for 25 min. Afterwards, their learning was tested and confirmed in a recognition test (all performed with >70% accuracy in recognizing correct vs. incorrect notes). Judges were given verbal descriptions of the four concepts (novelty, correctness, pleasantness, and creativity). Specifically, a composition was considered novel if it contained unique combinations of notes compared to the artificial grammar or to the compositions of other participants. The more correct intervals a composition contained the higher its correctness. Pleasantness constituted a subjective measure of enjoyableness. Creativity was defined as combining novelty and grammaticality. Importantly, raters were instructed to provide ratings taking into account the constraints of the AMG.

2.3.1. *Training sessions*

For the training sessions, participants were seated in front of a computer whose keyboard was adjusted to serve as a sound keyboard: different coloured stickers were put on keys A, D, G, J, L, ‘, and ENTER; red, orange, yellow, green, blue, pink, and brown, respectively. The cues were spatially positioned on the screen, i.e. lower notes on the right and higher on the left.

Before the first session only, participants listened to the whole 7-note scale going up three times, while, at the same time, the corresponding colours of each note were presented on screen in coloured circles (visual cues). Then, they were given 3 min to familiarize themselves with the keyboard. Afterwards, they completed a short note discrimination test. Specifically, they listened to intervals (pairs of notes), for which they

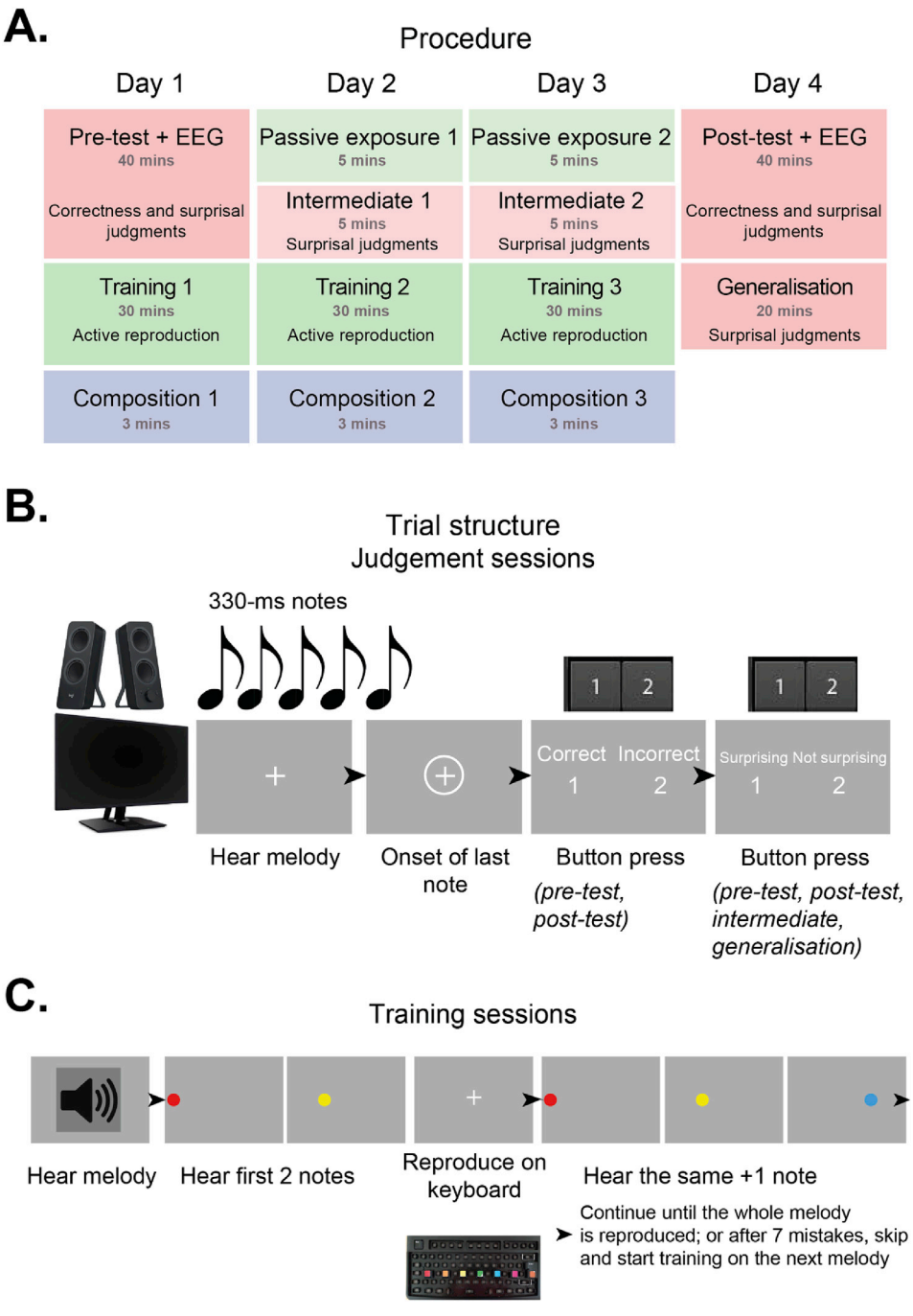


Fig. 2. A. An illustration of the experimental design; **B.** Illustration of trial structure of the judgement sessions. Participants heard a melodic sequence and were asked to judge if the last note was correct or incorrect (pre-test and post-test) and surprising or not surprising (pre-test, post-test, intermediate and generalisation sessions), by pressing 1 or 2 on a computer keyboard; **C.** Illustration of trial structure of the three training sessions. Participants listened to a melodic sequence. They then heard the first two notes and needed to reproduce them on the keyboard. If they reproduced them correctly, the next sequence would increase by one note, whereas if they made a mistake they would try again. After making 7 mistakes on one melody, they started training on the next melody.

were presented with only the first visual cue on screen, and were asked to reproduce the interval on the keyboard (i.e., they needed to identify and produce the second note). They were allowed three attempts, and, if unsuccessful, the solution was presented on screen at the end. There were 42 trials in total (covering all possible note combinations, e.g., C-D, C-E, C–F, etc). All participants had to pass an arbitrary threshold of 70% correct to proceed with the training.

There were three 25-min training sessions of active reproduction of the 12 old-grammatical sequences on the keyboard. Participants began by listening to a melodic sequence. Then the first 2 notes of the melody were presented, and only after participants reproduced them correctly, the next segment was increased by a note and so on. If they reproduced them incorrectly, the segment would repeat for further attempts. After 7 mistakes, the training on this melody was terminated, and they would start training on another melody. Half of the participants ($N = 19$) were presented with the visual cues of all the notes that they needed to reproduce on screen, while the other half ($N = 19$) were only given the first cue as a reference (to indicate the first note of the sequence), but relied only on the auditory information to reproduce the rest. Comparison of the two training methods is outside of the scope of this paper, and here we present findings after combining the two groups together. Both groups were tested and re-tested using the exact same method. Comparisons between the two groups revealed that the training method was associated with no difference in learning. The detailed analysis of the differences between these two groups is presented in Supplementary materials.

2.3.2. Test sessions

To assess learning of the AMG, test sessions were conducted before (pre-test: day 1) and after (post-test: day 4) the training. Participants were seated in front of the computer, while EEG was recorded. Through written instructions, they were informed that they would listen to melodies of a novel music grammar governed by a set of rules. They were instructed to attend as the melodies would stop at random points and they would be prompted to make two judgements on the last note: (i) correct or incorrect, and (ii) surprising or not surprising, by pressing keys 1 and 2 on a number pad, respectively. Three practice trials familiarised them with the task. Across participants, the presentation order of the trials was randomised. There were 280 trials in total. Three breaks were provided, and each session lasted around 40 min.

In the generalisation surprisal session, participants were presented with the new-grammatical sequences in randomized order, and were prompted to judge if the last note was surprising or not surprising. There were 105 trials in total. This session lasted around 20 min.

2.3.3. Passive exposure sessions

To ensure successful learning of the novel music grammar, on days 2 and 3, participants were exposed to three blocks of all old-grammatical sequences in randomised order following Rohrmeier et al. (2011). They were instructed to listen attentively to the melodies. There was a total of 36 trials and the session lasted around 5 min.

2.3.4. Intermediate surprisal sessions

Just after the exposure sessions, on days 2 and 3, participants were presented with sequences terminating on high-probability and low-probability notes and were asked to judge if the last note was surprising or not surprising. That was a total of 36 trials and lasted around 7 min.

2.3.5. Creativity sessions

After each training session (days 1–3) participants were asked to create one musical composition. They were given verbal instructions that they should create something based on what they learned, and also make it as creative as possible. For the preparation, they were given 3 min and they could play with the keyboard and use pen and paper. Then they were given 20 s to perform their composition on the keyboard while it was being recorded through MATLAB.

2.4. EEG recording and preprocessing

The EEG signals were recorded from 64 Ag–AgCl electrodes attached to the EGI geodesic sensor net system (HydroCel GSN 64 1.0; EGI System 200; Electrical Geodesic Inc., OR, USA; https://www.egi.com/) and amplified by an EGI Amp 300. The sampling frequency was 500 Hz, and the signals were high-pass filtered at 0.1 Hz. The MATLAB Toolbox EEGLAB (Delorme and Makeig, 2004) was used for data preprocessing, and FieldTrip (Oostenveld et al., 2011) for data analysis. EEG data were recorded with an online reference in the right mastoid and re-referenced to the average of the right and left mastoid electrodes. Continuous data were high-pass filtered at 0.5 Hz and then epoched from −0.1 to 0.6 s around the onset of the last note. Data from electrodes with consistently poor signal quality, as observed by visual inspection and by studying the topographical maps of their power spectra, were removed and replaced by interpolating neighbouring electrodes. Artefact rejection was conducted in a semi-automatic fashion. First, artefactual epochs were removed by visual inspection. Independent component analysis was used to correct for eye-blink related artefacts. The epoched data was low-pass filtered at 30 Hz and baseline corrected from −0.1 to 0 s.

2.5. Data analysis

2.5.1. Behavioural analysis

Performance during learning and training. Participants' level of learning was assessed during the test sessions, the intermediate exposure sessions, and the training sessions. Specifically, in the test sessions, a response was considered correct if a high-probability (HP) or low-probability (LP) note was judged as correct, and if an incorrect note (INC) was judged as incorrect.

To investigate participants' sensitivity to the statistical probabilities of the novel music grammar, we calculated the percentage of notes judged as surprising in the pre- and post-test within each note probability category. We performed a 3 (*note probability:* HP vs. LP vs. INC) x 2 (*session:* pre vs. post) repeated measures ANOVA. The same was applied to the two intermediate surprisal sessions, where a 2 (*note probability:* HP vs. LP) x 2 (*intermediate session:* 1, 2) repeated measures ANOVA was performed. A repeated measures ANOVA with *note probability* as the independent variable (HP vs. LP vs. INC) and surprisal judgement as the dependent variable was used to assess generalization in the final test session.

Further, we evaluated performance during the training sessions by calculating the mean length of correctly reproduced sequences (in number of notes). Due to technical problems with saving the results, four participants were excluded from this analysis ($N = 34$). A one-way repeated measures ANOVA with three levels (*training session*: 1, 2, 3) and sequence length as the dependent variable was conducted.

2.5.2. EEG analysis

ERP analysis. Five further participants were removed due to poor EEG data quality (more than 30% of the trials rejected in at least one of the test sessions) ($N = 33$). The following ERP components were analysed: the N100, the P200, and the P3a in fronto-central regions (E8, E6, E4 in the EGI configuration, corresponding to: AFz, Fz, FCz in the standard 10–20 system), and the P3b in parietal regions (E33, E36, E38 in the EGI configuration, corresponding to: P1, Pz, P2 in the standard 10–20 system). The peak amplitude of each participant was extracted from the N100 time window (70–150 ms), the P200 (170–250 ms), and the P3a and P3b (300–350 ms). A 2 x 2 x 2 factorial ANOVA was performed with the following factors: *accuracy* (correct response vs. incorrect response), *correctness* (correct note vs. incorrect note), and *session* (pre-test vs. post-test).

Time-frequency representation (TFR). To analyse oscillatory brain activity during the encoding and the retrieval phases of the AMG, we

conducted a time-frequency analysis time-locked to the onset of the final note of each melodic sequence of both the pre-test and the post-test sessions. We used responses to grammatical notes only (HP and LP), and excluded ungrammatical notes (INC and random). The EEG signal was convolved with a complex Morlet wavelet. The TFR was calculated from 1 to 70 Hz, in 50 logarithmic exponential steps, using 5-cycle wavelets. Then, the wavelet convolved values were averaged from −0.2 to 1 s time-locked to note onset. We divided the power after note onset by the average baseline power from −0.2 to 0 s (note onset). Finally, we transformed all the power values to their base 10 logarithms.

Encoding neural mechanisms: To explore potential relations between the pre-test TFRs and learning, we conducted Pearson's correlations between pre-test power and post-test improvement accuracy (post-test minus pre-test accuracy). As there was no solid hypothesis to justify a hypothesis-driven analysis and considering the multiple comparisons problem, we followed a cross-validation procedure for the statistical analysis. Specifically, we randomly split the EEG trials of each participant in two parts, part A and part B. All part A's (dataset A) were used to explore meaningful correlations and identify regions of interest (ROIs). The criteria to identify a meaningful correlation was that it should be more than 100 ms long. Subsequently, all part B's (dataset B) were used to validate the significant effects in the previously identified ROIs with meaningful correlations. The regions of interest identified were: left temporal (E24, E25, E19 in EGI; corresponding to T7, TP7, FT7 in the standard 10–20 system), fronto-central (Cz, E4, E6 in EGI; Cz, FCz, Fz in 10–20 system), left frontal (E13, E14, E15 in EGI; FC3, F5, F3 in 10–20 system), and right frontal (E58, E56, E53 in EGI; FC4, F6, F4 in 10–20 system). Participants were grouped into low-learners, LL ($N = 19$) and high-learners, HL ($N = 19$) using a median split based on their accuracy in the post-test. We then compared neural responses in HL vs. LL.

Retrieval neural mechanisms: Pearson's correlations between post-test power and post-test accuracy were conducted to investigate the neural signatures of retrieval of the novel material. We compared neural responses in HL vs. LL. Finally, the identified ROIs were correlated with the subjective measures (novelty, correctness, pleasantness, and creativity) of the participants' compositions. Further, we explored correlations between the ERP indices of learning in the post-test (N1–P2 peak-to-peak amplitude in response to incorrect notes) and post-test accuracy.

2.5.3. Analysis of the musical compositions of participants

In section 3.2 we present the main results of the analyses and mark all significant contrasts with asterisks in the respective figures. Planned contrasts are reported in detail in Supplementary materials.

Human judgements: Four expert judges evaluated the compositions on four aspects - novelty, correctness, pleasantness, and creativity - on a scale from 1 (not at all) to 5 (extremely). We observed a reasonable agreement between the four raters (interclass correlation, IC) for *novelty* (IC = 0.60; CI: 0.51-0.67), *correctness* (IC = 0.66; CI: 0.59-0.73), and *creativity* (IC = 0.56; CI: 0.47-0.65), and a reduced agreement on *pleasantness* (IC = 0.38; CI: 0.24-0.50). Z scores were calculated separately for each judge and aspect, and then averaged over judges. First, Pearson's correlations were conducted to investigate the relationship between the four measures. Subsequently, in order to track how these changed throughout training, we conducted 3 (*session*: 1, 2, 3) x 2 (*group:* low-learners vs. high-learners) mixed ANOVAs separately for each measure. Correlations between human judgements are presented in Supplementary materials.

Objective measures: The musical compositions of the participants were assessed using objective measures. First, we calculated the percentage of correct intervals in each composition as a measure of correctness or grammaticality. As described in the Materials (section 2.2), the AMG comprised of 9 intervals ('musical terminals', i.e. pairs of notes), which are combined to create different melodic sequences. In order to identify the number of correct intervals (i.e. note pairs) in each composition, we calculated how many of those 9 intervals of the AMG are encountered in each composition. To normalize the number of correct intervals based on the length of each composition, we then divided that by the total number of notes of the composition, which comprised a measure of the percentage of correct intervals. Second, we computed the length (in number of notes) of the compositions. Third, in order to investigate how participants used their knowledge of the grammar, we calculated the probabilities of the correct only notes, as well as of all the notes, using IDyOM (pretrained on the sequences that each participant heard previously). Using a high number of high-probability notes would show that participants produced compositions based on the learned AMG, whereas more low-probability notes would mean that they used more unpredictable or incorrect notes based on the grammar. Four separate 3 (session:1, 2, 3) x 2 (group: low-learners vs. high-learners) mixed ANOVAs were conducted with percentage of correct intervals, number of notes, IDyOM probability of the correct notes, and IDyOM probability of all notes as the dependent variables, respectively.

In Supplementary materials, we presented the results from Pearson *r* correlations between the subjective and objective measures of the musical compositions under the section "Correlations between subjective and objective measures". Furthermore, in Supplementary materials, we included additional analyses demonstrating that participants' musical compositions were based on their knowledge of the AMG rather than Western music ("Probabilities of music compositions according to the AMG vs. Western music"), and that the judgements of the human experts were not confounded by cultural familiarity with the Western style ("Correlations between IDyOM probabilities and human judgements").

2.5.4. Predicting creativity from computational and quantitative measures

To evaluate the predictive strength of the computational measures of the compositions for the creativity ratings of the experts, we constructed three regression models with perceived creativity as the dependent variable. First, we investigated whether learning (accuracy in the post-test) predicts creativity (rated by the judges). Second, based on previous literature identifying an inverted U-shaped relationship between liking and complexity (for a review see Chmiel and Schubert, 2017; Güçlütürk et al., 2016), we assessed quadratic relationships between the percentage of correct intervals and note probability of correct notes (estimated from IDyOM). Third, we investigated which of the previously identified brain measures predict creativity: N100, P200, and P3a amplitude in response to incorrect notes in the post-test, alpha power at Pz and beta power at T7 in the post-test, and delta power at T7 in the pre-test. We predicted the perceived creativity of session 3 only, as this took place after training when participants had complete knowledge of the AMG. We investigated both linear and quadratic relationships between the predictor variables and the dependent variable (i.e. creativity) in all models.

2.5.5. Control measures

Participants completed a working memory span task to control for the effect of working memory on performance at the training sessions. Further, we conducted intertrial phase coherence analysis in the pre- and post-test sessions, which measures the consistency of phase values at a given frequency and time point. This was done to rule out the possibility that the identified neural mechanisms of learning represented mere entrainment to the properties of the stimuli. More details and the results of both control analyses are in Supplementary materials.

3. Results

3.1. Behavioural results

3.1.1. Pre-test and post-test

Participants showed a significant improvement in their accuracy in recognizing grammatical notes from pre-to post-test as confirmed by a paired *t*-test ($t(37) = 8.339$, $p < .001$, *Cohen's d* = 1.353) (Fig. 3A).

Participants judged low probability and incorrect notes as more surprising than notes of high probability, evidencing that learning made them more sensitive to the statistical probabilities of the AMG (Fig. 3B). A

3 (*note probability:* high-probability vs. low-probability vs. incorrect) x 2 (*session:* pre vs. post) repeated measures ANOVA on the percentage of notes judged as surprising revealed significant main effects of *note probability* ($F(2,74) = 46.566$, $p < .001$, $\eta^2 = 0.557$), as well as a *note probability* * *session* interaction ($F(2,74) = 28.719$, $p < .001$, $\eta^2 = 0.437$). Planned contrasts revealed that participants judged significantly fewer high- (HP) and low-probability (LP) notes as surprising in the post-test compared to the pre-test (HP: $t(37) = -3.982$, $p < .001$, *Cohen's* $d = -0.650$; LP: $t(37) = -2.841$, $p = .007$, *Cohen's* $d = -0.472$), whereas the opposite was found for the incorrect (INC) notes, i.e. participants judged them more as surprising in the post-session ($t(37) = 3.331$, $p = .002$, *Cohen's* $d = 0.559$). Further, in both sessions, INC were judged as surprising significantly more often than HP notes (pre: $t(37) = 3.913$, $p < .001$, *Cohen's* $d = 0.635$; post: $t(37) = 7.108$, $p < .001$, *Cohen's* $d = 1.159$), as well as LP notes (pre: $t(37) = 2.623$, $p = .013$, *Cohen's* $d = 0.428$; post: $t(37) = 6.741$, $p < .001$, *Cohen's* $d = 1.102$). The LP notes were also judged significantly more surprising than the HP notes (pre: $t(37) = 2.362$, $p = .024$, *Cohen's* $d = 0.387$; post: $t(37) = 4.926$, $p < .001$, *Cohen's* $d = 0.801$).

A paired *t*-test between the percentage of random notes judged as correct in the post-compared to the pre-test session confirmed that the effects are not due to a general bias towards judging notes as correct in the post-test ($t(37) = 1.219$, $p = .230$, *Cohen's* $d = 0.198$).

3.1.2. Intermediate sessions

Participants' surprisal judgements in the two intermediate sessions were also evaluated by a 2 (*note probability:* HP vs. LP) x 2 (*intermediate session:* 1, 2) repeated measures ANOVA (Fig. 3C). Results revealed significant main effects of *session* ($F(1,36) = 4.860$, $p = .034$, $\eta^2 = 0.119$) and *note probability* ($F(1,36) = 5.135$, $p = .030$, $\eta^2 = 0.125$). There was also a significant *note probability* * *session* interaction ($F(1,36) = 49.013$, $p < .001$, $\eta^2 = 0.577$). Planned contrasts revealed that participants judged HP notes as significantly less surprising in the second compared to the first session ($t(37) = -6.741$, $p < .001$, *Cohen's* $d = -1.094$), whereas the opposite was found for the LP notes, i.e. participants judged them as more surprising in the second session ($t(37) = 3.411$, $p = .002$, *Cohen's*

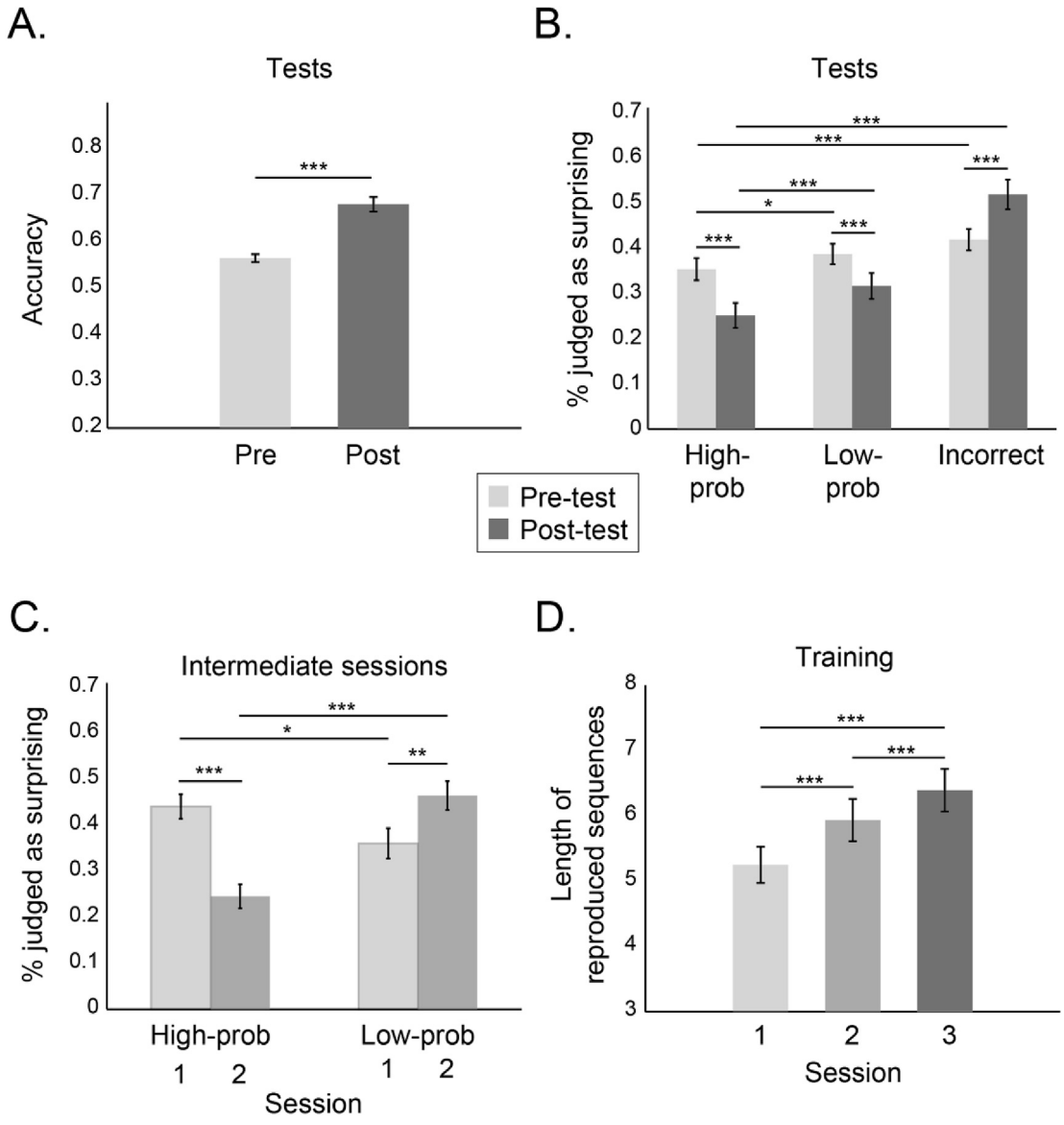


Fig. 3. Performance on the tests and training sessions. **A.** Mean accuracy in the pre-test (light grey) and post-test (dark grey) sessions; **B.** Mean percentage judged as surprising in the pre-test and post-tests separately for high-probability, low-probability, and incorrect notes; **C.** Mean percentage of notes judged as surprising in the two intermediate sessions separately for high-probability and low-probability notes; and **D.** Mean length (number of notes) of correctly reproduced sequences across the three training sessions. Error bars represent ±1 standard error mean (*SEM*). *$p < .050$, **$p < .010$, and ***$p < .001$.

$d = 0.554$). Further, in the first session, HP notes were judged as surprising significantly more often than the LP ones ($t(37) = -2.080$, $p = .045$, *Cohen's* $d = -0.340$), whereas the opposite effect was observed in the second session ($t(37) = 6.103$, $p < .001$, *Cohen's* $d = 0.996$).

3.1.3. Generalization test

A repeated measures ANOVA with *note probability* as the independent variable (HP vs. LP vs. INC) and surprisal judgement as the dependent variable in the generalization session demonstrated that participants successfully differentiated between the statistical probabilities of novel sequences which belonged to the grammar but were not heard during learning (main effect of *note probability:* $F(2,74) = 42.301$, $p < .001$, $\eta^2 = 0.533$). Planned contrasts showed that participants judged LP notes as more surprising than HP ones ($t(37) = 6.039$, $p < .001$, *Cohen's* $d = 0.980$), and INC notes more surprising than both HP ($t(37) = 7.616$, $p < .001$, *Cohen's* $d = 1.235$) and LP ones ($t(37) = 4.970$, $p < .001$, *Cohen's* $d = 0.806$).

3.1.4. Training

Performance during the training sessions improved incrementally, as confirmed by a one-way ANOVA with mean length of reproduced sequences as the dependent variable (Fig. 3D). Participants managed to correctly reproduce an increasing number of notes throughout the sessions (main effect of *session:* $F(2,66) = 38.012$, $p < .001$, $\eta^2 = 0.535$). Paired *t*-tests confirmed that participants performed better throughout the sessions (second-first: $t(33) = 4.740$, $p < .001$, *Cohen's* $d = 0.813$; third-second: $t(33) = 5.918$, $p < .001$, *Cohen's* $d = 1.015$); and third-first: $t(33) = 7.181$, $p < .001$, *Cohen's* $d = 1.231$).

3.2. ERP results

3.2.1. N100 time window (70–150 ms)

Incorrect notes elicited higher N100 amplitude compared to correct notes, both in the pre- and post-test, but only when participants responded correctly (Fig. 4). A 2 (*accuracy:* correct response vs. incorrect response) x 2 (*correctness:* correct note vs. incorrect note) x 2 (*session:* pre vs. post) factorial ANOVA revealed an *accuracy* * *correctness* interaction ($F(1,32) = 7.652$, $p = .009$, $\eta^2 = 0.193$). Planned contrasts showed that this effect was due to incorrect notes eliciting significantly higher N100 compared to correct notes for correct responses ($t(32) = -4.159$, $p < .001$, *Cohen's* $d = 0.724$), but not for incorrect responses ($t(32) = 0.725$, $p = .474$, *Cohen's* $d = 0.126$).

3.2.2. P200 time window (170–250 ms)

As with the N100, in the P200 time window incorrect notes elicited higher amplitudes compared to correct notes during correct only

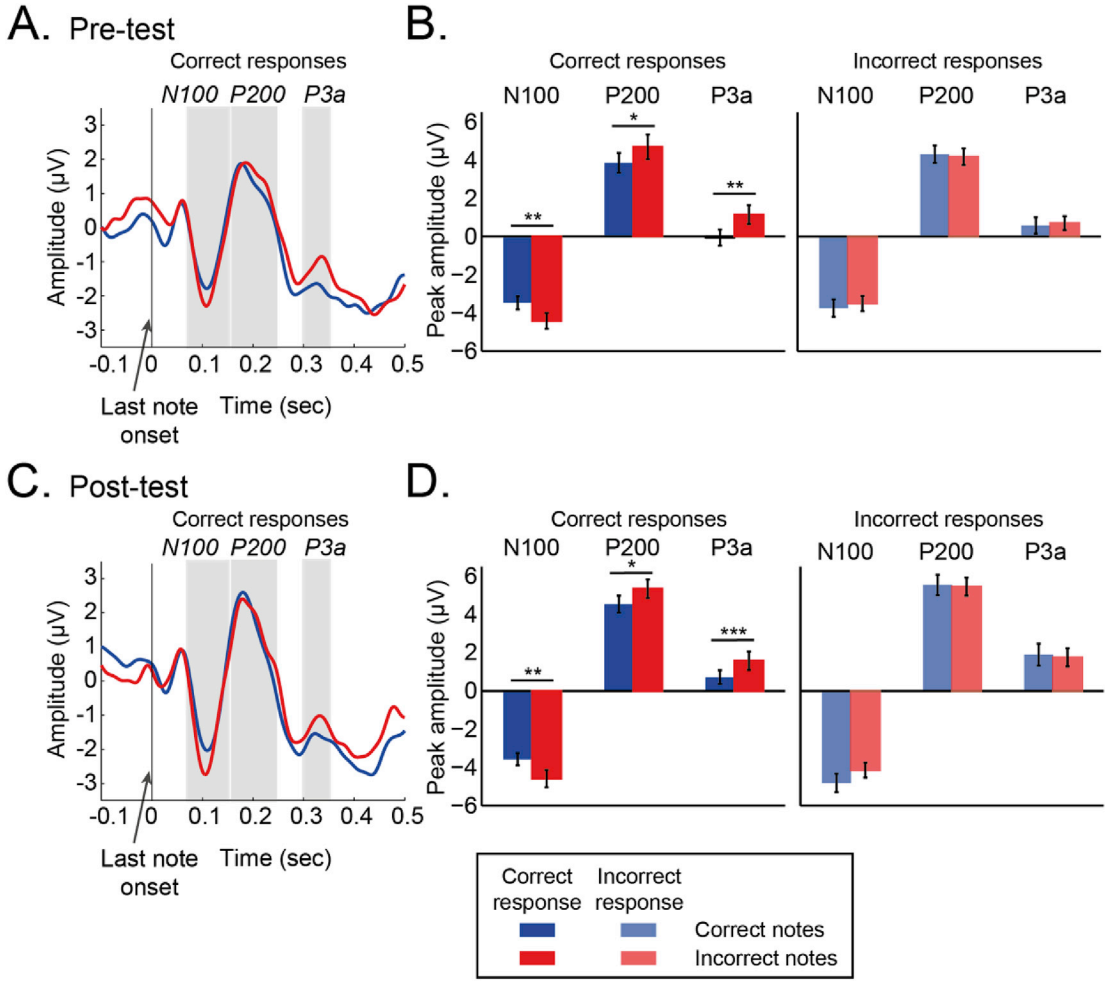


Fig. 4. Grand average ERPs averaged over fronto-central brain regions (average over electrodes AFz, Fz, FCz) in response to the last note of the melodic sequences during the pre-test **(A.)** and post-test **(C.)** sessions, for participants' correct responses. The blue line represents correct notes, whereas the red line represents incorrect notes. The shaded rectangles indicate the N100 (70–150 ms), the P200 (170–250 ms), and the P3a (300–350 ms) time windows; **B.** Mean peak N100, P200, and P3a amplitudes in response to correct (blue) and incorrect (red) notes, for participants' correct (opaque) and incorrect (transparent) responses in the pre-test; **D.** Same as B. but for post-test. Error bars represent ± 1 standard error mean (*SEM*). *$p < .05$, **$p < .01$, ***$p < .001$.

responses (Fig. 4). An ANOVA revealed a marginal *accuracy * correctness* interaction ($F(1,32) = 4.524$, $p = .041$, $\eta^2 = 0.124$). Planned contrasts showed that this effect was due to incorrect notes eliciting significantly higher P200 compared to correct notes for correct responses ($t(32) = 2.990$, $p = .005$, *Cohen's* $d = 0.520$), but not for incorrect responses ($t(32) = -0.310$, $p = .758$, *Cohen's* $d = 0.054$). There was also a main effect of session ($F(1,32) = 4.860$, $p = .035$, $\eta^2 = 0.132$), as P200 was higher in the post-test ($M = 4.918$) compared to the pre-test ($M = 4.251$).

3.2.3. Fronto-central P3a time window (300–350 ms)

As with the N100 and the P200 time window, incorrect notes elicited higher P3a amplitudes compared to correct notes during correct only responses (Fig. 4). An ANOVA revealed a marginal *accuracy * correctness* interaction ($F(1,32) = 5.578$, $p = .024$, $\eta^2 = 0.148$). Planned contrasts showed that this effect was due to incorrect notes eliciting significantly

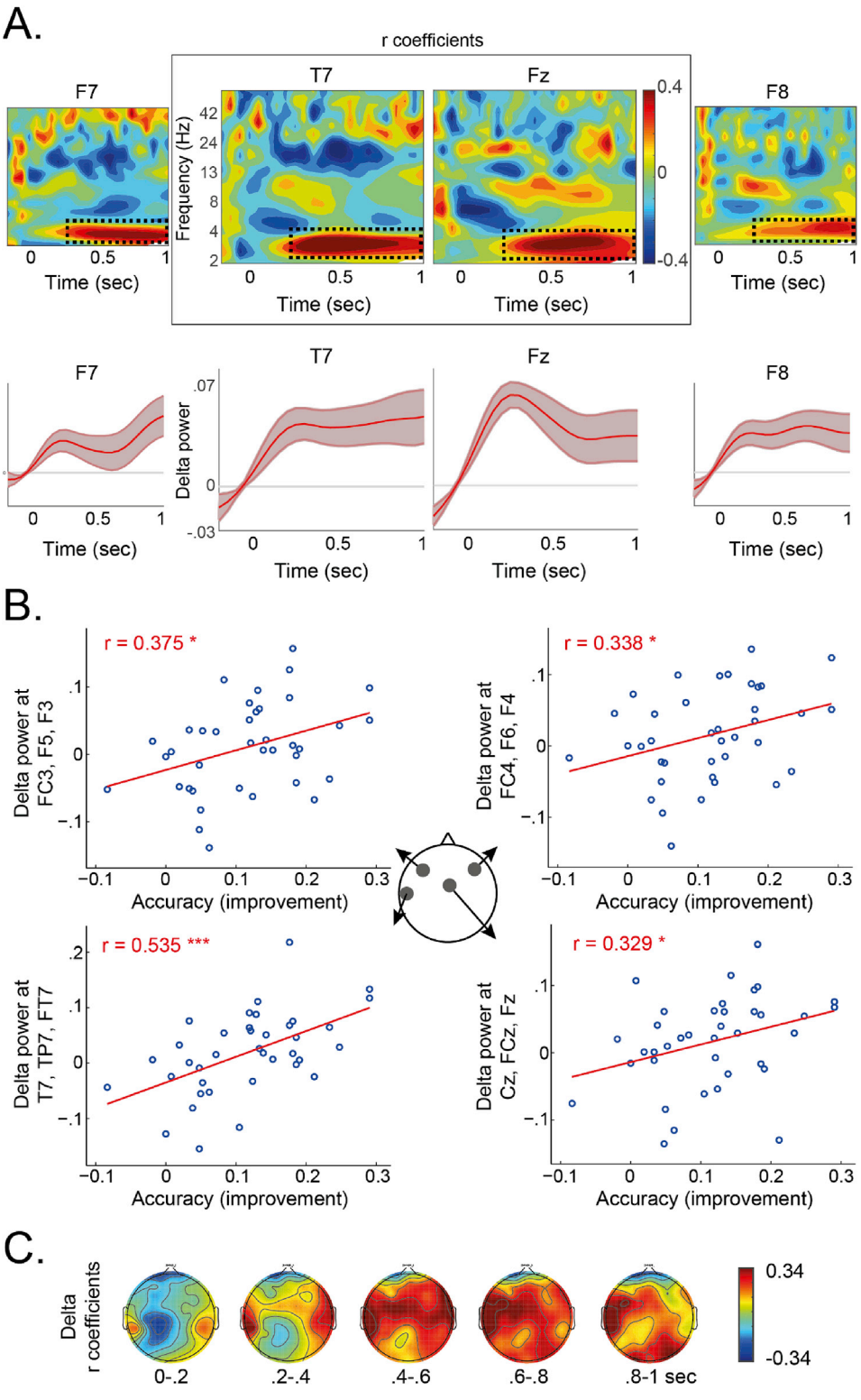


Fig. 5. Correlations between delta band power in the pre-test and accuracy improvement. **A.** Top: The plots show the Pearson's coefficients of correlations between pre-test EEG power and post-test accuracy improvement, separately for electrodes T7, Fz, F7, and F8. The marked regions in squares designate the selected regions of interest. Bottom: Time course of pre-test power over the delta (2.5–4.5 Hz) frequency band for each electrode of interest; **B.** Pearson's correlations between pre-test delta (2.5–4.5 Hz) power and post-test accuracy improvement, separately for left temporal (T7, TP7, FT7), fronto-central (Cz, FCz, Fz), left frontal (FC3, F5, F3), and right frontal (FC4, F6, F4) regions in the cross-validation dataset. Statistically significant *r* correlation coefficients are marked in red, with: $*p < .05$, $**p < .01$. To confirm the validity of the results in the presence of the outliers observed in the figure, we conducted Spearman's ρ correlations (delta in left temporal: Spearman's $\rho = 0.500$, $p < .001$; left frontal: Spearman's $\rho = 0.352$, $p = .030$; fronto-central: Spearman's $\rho = 0.339$, $p = .037$; right frontal: Spearman's $\rho = 0.346$, $p = .034$); **C.** Topographies of *r* coefficients for correlations between pre-test delta power and post-test accuracy improvement.

higher P3a compared to correct notes for correct responses ($t(32) = 5.462$, $p < .001$, *Cohen's* $d = 0.951$), but not for incorrect responses ($t(32) = -0.052$, $p = .959$, *Cohen's* $d = 0.009$). There was also a main effect of correctness ($F(1,32) = 7.426$, $p = .010$, $\eta^2 = 0.188$), as P200 was higher in the incorrect notes ($M = 0.911$) compared to the correct notes ($M = 0.305$).

3.2.4. Parietal P3b time window (300–350 ms)

An ANOVA revealed no effect or interaction between the variables on the P3b ($p > .09$).

3.3. Neural mechanisms of encoding unfamiliar melodic sequences

As stated in the Methods section 2.5.2, only grammatical (HP and LP notes) trials were used for the TFR analysis. We first explored correlations between pre-test TFR and post-test accuracy improvement in dataset A. Subsequently, we conducted statistical analyses in dataset B. This allowed to avoid circularity in our analysis by testing the observed correlations in a new set of data. In Fig. 5A (top), Pearson's correlation coefficients in electrodes T7, Fz, F7, and F8 are plotted. We observed statistically significant positive correlations in the delta frequency band (2.5–4.5 Hz) around 0.25–1 s post note onset.

3.3.1. Cross-validation of correlations

In order to validate the observations in dataset A, we used the unseen dataset B to conduct the correlations in the selected ROIs (Fig. 5B and C). Specifically, we carried out correlations between improvement accuracy and pre-test power in delta band in left temporal (T7, TP7, FT7), fronto-central (Cz, FCz, Fz), left frontal (FC3, F5, F3), and right frontal (FC4, F6, F4), from 0.25 to 1 s post note onset. There was a significant positive correlation between pre-test delta power and improvement in all four regions: left temporal ($r = 0.535$, $p < .001$), left frontal ($r = 0.375$, $p = .020$), fronto-central ($r = 0.330$, $p = .044$), and right frontal ($r = 0.339$, $p = .038$) regions. As suggested by these cross-validation results, it seems that the higher the power in the delta frequency band in the pre-test in response to grammatical notes, the higher the improvement in post-test accuracy in learning of the novel music grammar. This effect was observed in left temporal areas, as well as in more central and frontal bilateral sites, suggesting a potential link between delta band oscillations and encoding of novel sequential material.

3.3.2. High-learners vs. low-learners

If delta activity represents an encoding mechanism, we expected that high-learners would exhibit higher delta power in the pre-test in response to the grammatical notes, compared to low-learners, and that this effect would disappear in the post-test (Fig. 6A and B). This was confirmed by independent samples *t*-tests comparing high-vs. low-learners in the pre-test for F7 ($t(35) = 2.109$, $p = .042$), T7 ($t(35) = 2.511$, $p = .017$), Fz ($t(35) = 2.367$, $p = .023$), and F8 ($t(35) = 2.688$, $p = .011$). No significant difference was found between groups in the post-test ($p > .155$).

3.3.3. Delta power of remembered vs. not remembered sequences

Here we define as "remembered sequences" the sequences which were judged as correct in the post-test (i.e. were successfully encoded in the pre-test and remembered in the post-test), while "not remembered sequences" are the sequences which were incorrectly judged in the post-test. To examine the effect of delta oscillations on the trial level, we calculated the delta power in response to test notes during the pre-test, which were successfully remembered (i.e. judged as correct) in the post-test vs. the notes which were not remembered (i.e. judged as incorrect) in the post-test. The *t*-value topographic map (Fig. 6C) revealed that the difference was located in central areas. Paired *t*-tests between remembered vs. not remembered at Cz revealed a significant difference in the pre-test values ($t(35) = 2.051$, $p = .044$), but not in the post-test ($p = .777$).

3.4. Neural mechanisms of retrieving learned material

To investigate neural mechanisms of retrieval of the learned material, we performed Pearson's correlations between post-test TFR and post-test accuracy. We observed a meaningful positive correlation in the upper beta frequency band (18–32 Hz, from now on "beta") at T7, as well as a negative correlation in the upper alpha band (10–13 Hz) at Pz from 0.15 to 1 s post note onset. Post-test beta power and post-test accuracy in T7 from 0.15 to 1 s were found to be significantly correlated (Pearson's $r = 0.521$, $p = .001$; Spearman's $\rho = 0.464$, $p = .004$). After exclusion of four outliers the correlation was still significant (Pearson's $r = 0.417$, $p = .016$) (Fig. 7A). Post-test alpha power and post-test accuracy in Pz were also significantly correlated ($r = 0.390$, $p = .017$) (Fig. 7B).

3.4.1. High-learners vs. low-learners

As post-test left temporal beta and parietal alpha band activity were found to be correlated with post-test accuracy, we expected power differences between high- and low-learners in the post-test but not in the pre-test. This was confirmed by independent samples *t*-tests comparing high-vs. low-learners in the post-test on T7 beta ($t(35) = 2.326$, $p = .026$), and Pz alpha power ($t(35) = -2.503$, $p = .017$) (Fig. 8A and B). No significant group differences were found in the pre-test ($p > .109$).

3.5. Analysis of the melodic compositions of participants

3.5.1. Human judgements throughout the sessions for high-learners (HL) vs. low-learners (LL)

Creativity: There was a significant *session* * *group* interaction ($F(2,72) = 3.261$, $p = .044$, $\eta^2 = 0.083$) (top left Fig. 9A). HL composed more creative melodies compared to LL in the third session, as confirmed by planned contrasts (independent samples *t*-tests) ($t(36) = 2.193$, $p = .035$, *Cohen's* $d = 0.720$), however there was no significant difference in any of the other sessions ($p > .490$). Further, paired *t*-tests showed that HL produced more creative compositions in the third session compared to the first ($t(17) = 2.780$, $p = .013$, *Cohen's* $d = 0.803$) and in the third session compared to the second ($t(17) = 2.552$, $p = .021$, *Cohen's* $d = 0.777$). There were no significant differences between any other contrast ($p > .443$). Neither the main effect of session nor group was significant ($p > .232$).

To further investigate the relationship between learning and creativity, we conducted a correlation between learning accuracy in the post-test and creativity in trial 3 (after training was completed). Results showed that learning is positively correlated with creativity ($r = 0.431$, $p = .007$).

Novelty: In order to investigate whether participants created more unusual compositions throughout the sessions, we conducted a 3 x 2 mixed ANOVA on the novelty judgements (top right Fig. 9A). HL produced overall less novel compositions than LL (main effect of *group*: $F(1,36) = 5.301$, $p = .027$, $\eta^2 = 0.128$). There was no significant effect of session or interaction between the factors ($p > .200$).

Correctness: The judges considered that participants composed increasingly more grammatical melodies as training progressed (main effect of *session*: $F(2,72) = 7.776$, $p = .001$, $\eta^2 = 0.178$) (bottom left Fig. 9A). HL produced overall more grammatical melodies compared to LL (main effect of *group*: $F(1,36) = 6.607$, $p = .014$, $\eta^2 = 0.155$). There was no significant interaction between the factors ($p = .992$).

Pleasantness: There was a main effect of *session* ($F(2,72) = 3.582$, $p = .033$, $\eta^2 = 0.091$) as the pleasantness was higher in the later sessions for both groups (bottom right Fig. 9A). Post-hoc tests showed that compositions of HL were judged as more pleasant in the third session compared to the second ($t(17) = 2.647$, $p = .017$, *Cohen's* $d = 0.624$), but there was no significant difference in any other contrast ($p > .067$). There was no significant effect of group or interaction between the factors ($p > .305$).

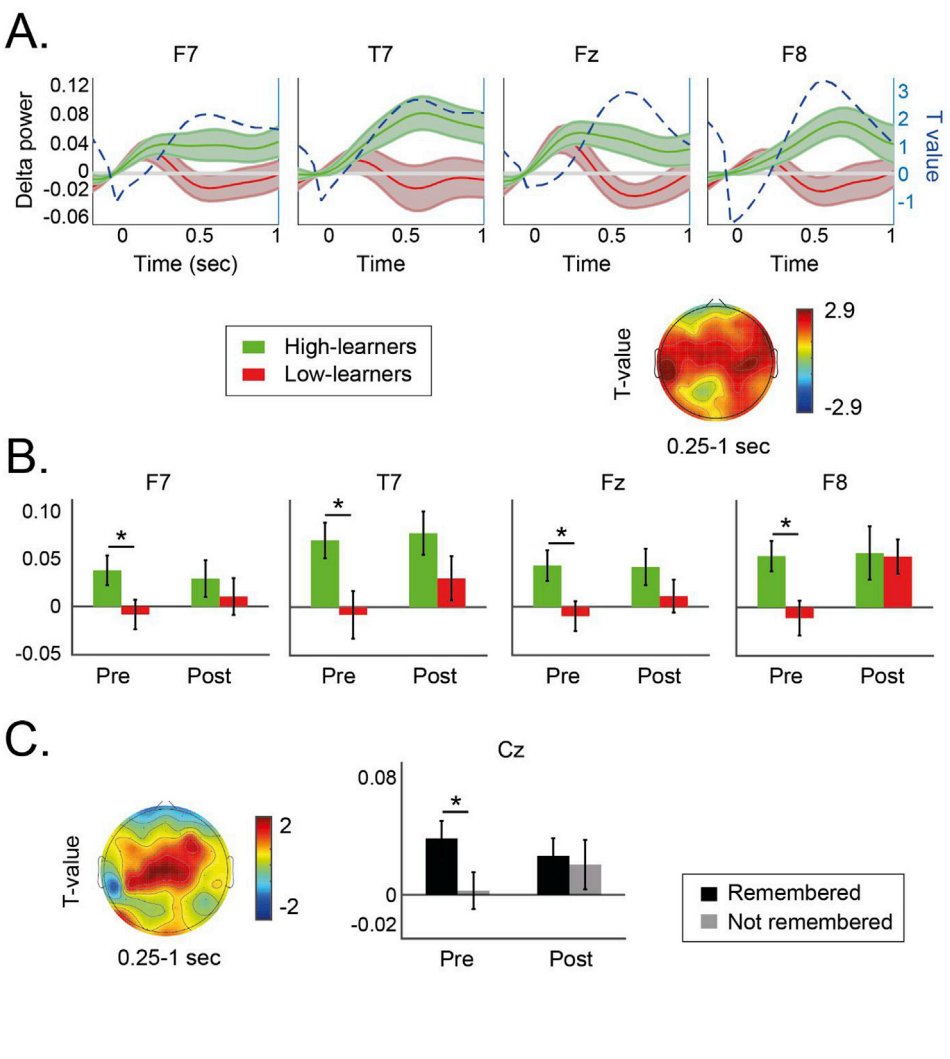


Fig. 6. A. Pre-test delta power in high-learners vs. low-learners. Top: Time course of delta power in F7, T7, Fz, and F8, separately for high- (green) and low- (red) learners, and the respective t-value (dashed blue) from timepoint-by-timepoint independent samples *t*-tests. Bottom: T-value topography of high-vs.- low-learners, averaged over 0.25–1 s post note onset; **B.** Mean delta power in pre-test (left) vs. post-test (right) for the electrodes and time window of interest, for high- (green) and low- (red) learners; **C.** Pre-test delta power of remembered (correctly judged in the post-test) vs. not remembered (incorrectly judged) target notes. Left: T-value topography of delta power averaged over 0.25–1 s post note onset, in response to remembered vs. not remembered notes. Right: Mean delta power for remembered (black) vs. not remembered (grey) sequences on Cz electrode, separately for pre-test and post-test. Error bars represent ± 1 standard error mean (*SEM*). $*p < .050$, $**p < .01$.

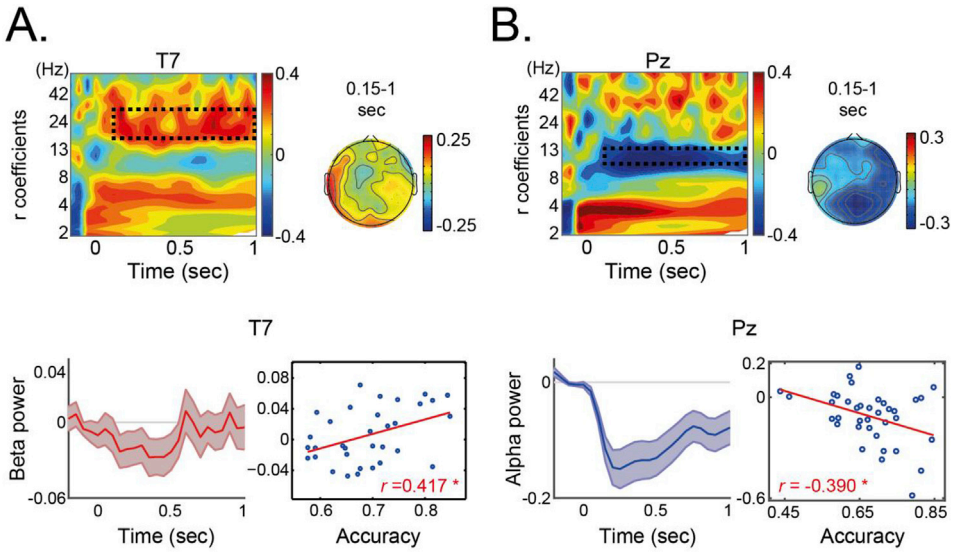


Fig. 7. A. Top left: Pearson's coefficients of correlations between post-test EEG power and post-test accuracy for T7. The marked region designates the selected region of interest (ROI: high beta/low gamma band: 18–32 Hz, 0.15–1 s). Top right: Topography of *r* coefficients averaged over 0.15–1 s post note onset for the beta band. Bottom left: Time course of post-test beta power in T7. Bottom right: Pearson's correlation between the beta ROI and post-test accuracy. **B.** The same as A. but for the alpha band (10–13 Hz) in Pz. $*p < .05$, $***p \leq .001$. To confirm the validity of the results, we conducted Spearman's ρ correlations (beta in T7: Spearman's $\rho = 0.464$, $p = .004$; alpha in Pz: Spearman's $\rho = -0.383$, $p = .019$).

3.5.2. Computational measures of compositions for high-learners (HL) vs. low-learners (LL)

Percentage of correct intervals: In order to investigate whether learning of the AMG was reflected in the compositions, we calculated the percentage of correct intervals across the three sessions (top left Fig. 9B). A 3 (*session:* 1, 2, 3) x 2 (*group:* LL vs. HL) mixed ANOVA was conducted on the percentage of correct intervals. Results showed that participants incorporated an increasingly higher number of correct intervals in their compositions as training progressed (main effect of *session*: $F(2,72) = 15.455$, $p < .001$, $\eta^2 = 0.300$). As expected, HL produced more correct intervals overall (main effect of *group*: $F(1,36) = 12.742$, $p = .001$, $\eta^2 = 0.261$). The interaction between session and group was not

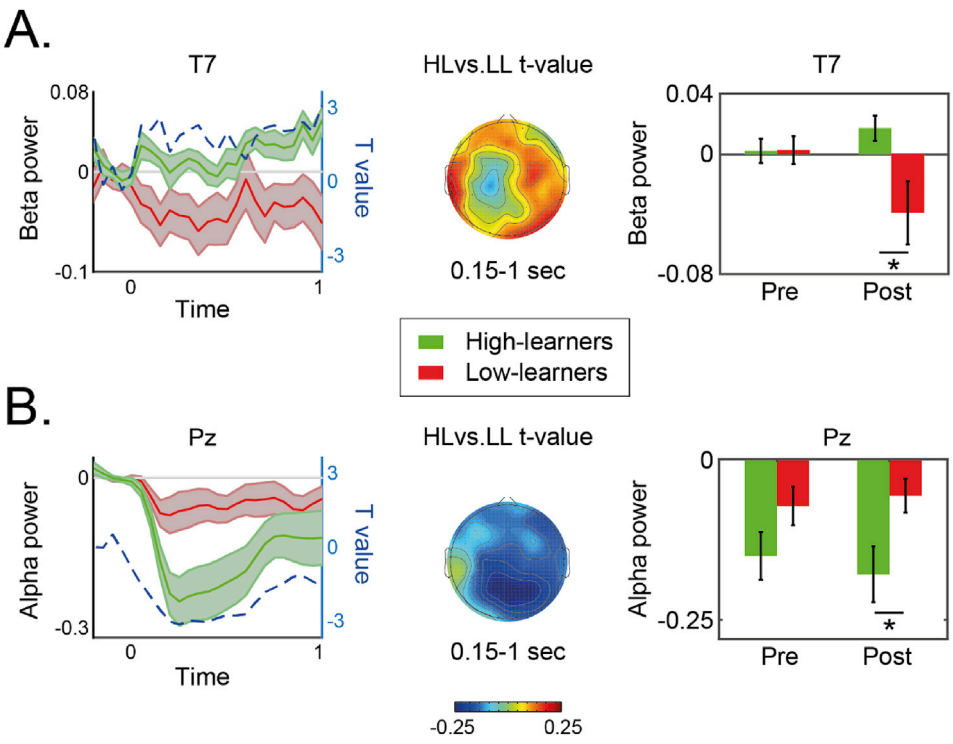


Fig. 8. A. Left: Time course of beta power in high-learners (green) vs. low-learners (red) in T7, and the respective *t* value (dashed blue). Middle: T-value topography between beta power averaged over 0.15–1 s post note onset of high-vs.- low-learners. Right: Mean beta power in pre-test vs. post-test, separately for each group; **B.** Same as A. but with alpha power. Error bars represent ± 1 standard error mean (*SEM*). *$p < .05$.

significant ($p = .390$).

Length of compositions: Participants created compositions with an increasing number of notes throughout the sessions (top right Fig. 9B), as confirmed by a 3 (*session*) x 2 (*group*) mixed ANOVA (main effect of *session*: $F(2,72) = 4.237$, $p = .018$, $\eta^2 = 0.105$). There was also a significant *session* * *group* interaction ($F(2,72) = 3.139$, $p = .049$, $\eta^2 = 0.080$). The effect of group was not significant ($p = .993$).

IDyOM probability of correct notes: Two participants were excluded from this analysis ($N = 36$) as they did not produce any correct notes. Participants used correct notes with increasingly higher probability throughout the sessions (bottom left Fig. 9B), as confirmed by a mixed ANOVA (main effect of *session*: $F(2,68) = 6.334$, $p = .003$, $\eta^2 = 0.157$). HL produced higher probability correct notes overall (main effect of *group*: $F(1,34) = 7.349$, $p = .010$, $\eta^2 = 0.178$). There was no session * group interaction ($p = .627$).

IDyOM probability of all notes: Participants incorporated notes with increasingly higher probability throughout the sessions (bottom right Fig. 9B), as confirmed by a mixed ANOVA (main effect of *session*: $F(2,72) = 10.459$, $p < .001$, $\eta^2 = 0.225$). HL produced higher probability notes overall (main effect of *group*: $F(1,36) = 7.136$, $p = .011$, $\eta^2 = 0.165$). There was no session * group interaction ($p = .260$).

3.5.3. Predicting creativity from quantitative measures

We investigated which quantitative features (neural correlates of learning, learning accuracy, computational features of compositions) predict more creative musical compositions as judged by human experts. First, we investigated whether learning predicts creativity. Second, we assessed whether the percentage of correct intervals and/or note probability of correct notes (as estimated from IDyOM) predicted creativity. Third, we investigated whether brain measures associated with learning can also predict creativity. Finally, we used the previously identified significant predictors to construct a final model predicting creativity from behavioural and neural features. We predicted the perceived creativity of session 3 only, as this took place after training when participants had complete knowledge of the AMG, and investigated both linear and quadratic relationships. Please refer to Supplementary materials for a report of correlations between the predictor variables used for the regressions.

3.5.3.1. Predicting creativity from learning. First, we created a regression model with accuracy in the post-test as predictor and creativity as the dependent variable. Learning was a significant linear predictor of creativity ($\beta = 1.490$, $p = .013$), and the model showed a significant overall fit of: $R^2 = 0.189$, $p = .043$. Learning was not related to creativity in a quadratic way ($\beta = 0.219$, $p = .704$).

3.5.3.2. Predicting creativity from computational measures. To evaluate the predictive strength of the computational measures of the compositions for the creativity ratings of the experts, we created a regression model with percentage of correct intervals and probability of correct only notes as predictors and creativity as the dependent variable. Percentage of correct intervals was a significant predictor of creativity with a quadratic trend ($\beta = -2.181$, $p < .001$), but there was no significant linear or quadratic relationship with the probability of correct notes ($p > .2$). This model was significant overall: $F(4,28) = 4.045$, $p = .010$, $R^2 = 0.366$.

3.5.3.3. Predicting creativity from EEG measures. A regression model was constructed with N100, P200, P3a amplitude in response to incorrect notes in the post-test, alpha power at Pz and beta power at T7 in the post-test, and delta power at T7 in the pre-test as predictors for creativity. Results revealed the P200 as a significant negative predictor of creativity ($\beta = -2.614$, $p = .011$). Neither of the other EEG predictors were significant ($p > .08$) nor the model was significant overall ($R^2 = 0.406$, $p = .386$).

3.5.3.4. Evaluation of predictors of creativity. In order to evaluate the contribution of each of the identified significant predictors (learning, percentage of correct intervals^2, and P200) on creativity, we

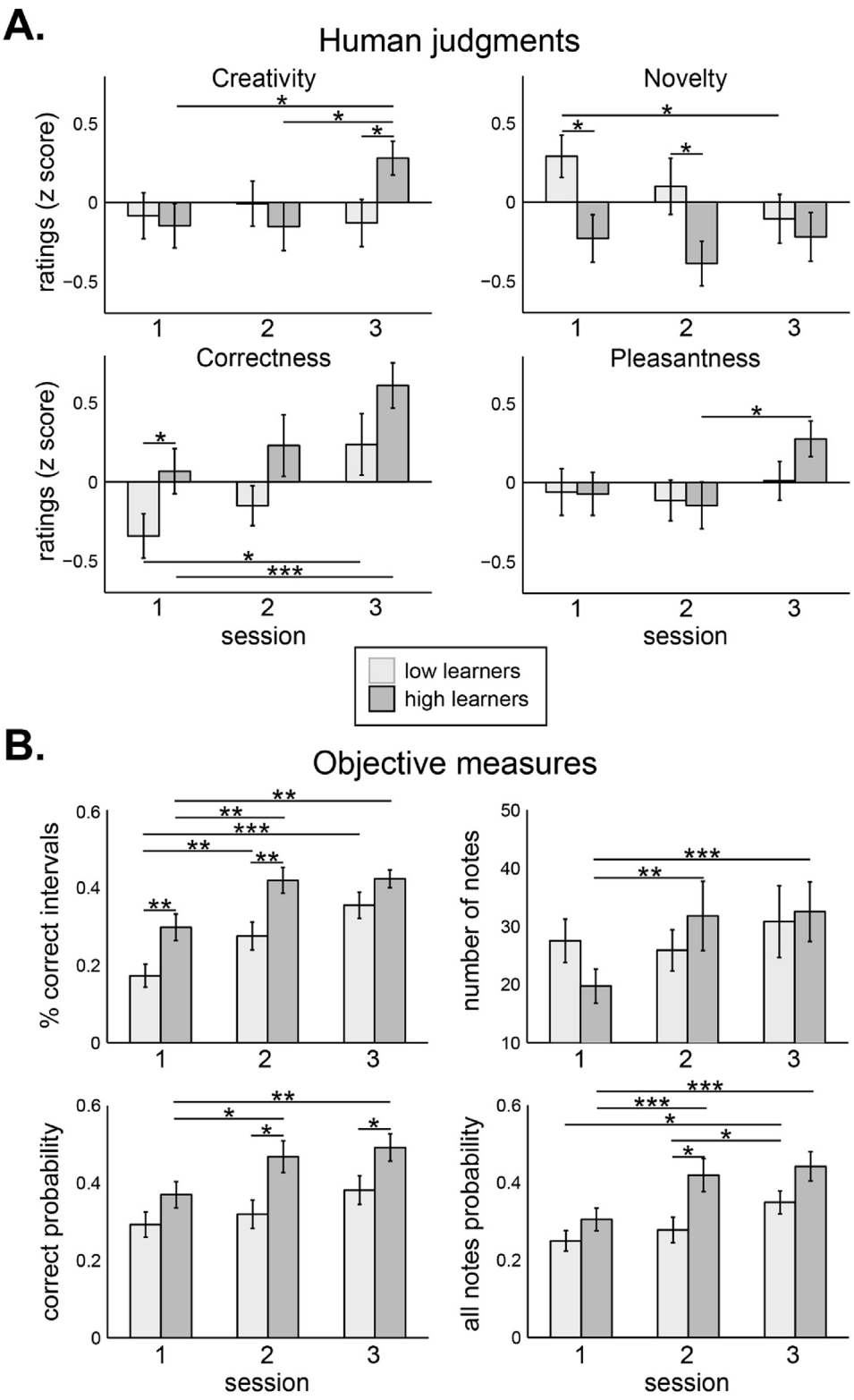


Fig. 9. A. Mean z-scored ratings of human judgements on: creativity, novelty, correctness, and pleasantness; **B.** Objective measures of grammaticality for participants' melodic compositions after 3 sessions of training on the artificial music grammar: percentage of correct intervals, number of notes, IDyOM probability of correct notes, and IDyOM probability of all notes. Error bars represent ±1 standard error mean (*SEM*). *$p < .050$, **$p < .010$, and ***$p < .001$.

constructed and compared three models. We incrementally compared fits by adding one predictor at a time. Specifically, Model 1 comprised a linear effect of learning; Model 2 was constructed from Model 1 by adding a quadratic effect of percentage of correct intervals; Model 3 included also a linear term of the P200 component in response to incorrect notes.

The above described models were compared using an ANOVA (see Table 1 below). Adding percentage of correct intervals (^{2}) to Model 1 led to a significantly improved fit over Model 1 ($p = .008$), while adding the P200 term further improved fit over Model 2 ($p = .004$). Model 3

including all predictors was also significantly better than Model 1 with accuracy only as predictor ($p = .001$). This is in line with the Residual Sum of Squares (RSS) values, showing substantial support for the above-mentioned terms.

Taking Model 3 as the final model, the regression coefficients indicate that accuracy affected creativity, as accuracy linearly increased creativity by about 1.029 ($t = 2.205$, $SE = 0.467$, $p = .035$) (Fig. 10, left). The percentage of correct intervals (quadratic term) affected creativity by −1.533 ($t = -3.285$, $SE = 0.467$, $p = .002$) (Fig. 10, middle). Further, the P200 amplitude in response to incorrect notes (linear term) decreased creativity by −1.405 ($t = -3.099$, $SE = 0.453$, $p = .004$) (Fig. 10, right). The final model (Model 3) was significant overall: $F(4,28) = 8.104$, $p < .001$, $R^2 = 0.537$.

3.5.3.5. Control analyses. The possibility that training on the AMG required only short-term memory skills was eliminated by control analysis performed on the working memory task. Further, intertrial phase coherence analysis in the pre- and post-test sessions controlled for the possibility that the identified neural mechanisms of encoding (delta band) and retrieval (alpha and beta bands) reflected mere entrainment processes. Please refer to Supplementary materials for a detailed report.

4. Discussion

In this study, we investigated the relationship between the behavioural and neural signatures of learning and creativity using an artificial music grammar (AMG). In brief, results showed that participants successfully learned an unfamiliar music grammar, as assessed by the test sessions and reflected in their ERP responses. The delta band power during first exposure to the unfamiliar melodies was positively correlated with accuracy improvement, suggesting this as a potential neural mechanism of encoding. On the other hand, retrieval mechanisms were associated with lower alpha and higher beta power after training.

Our results confirmed the crucial role of learning for creativity, as the level of learning robustly predicted subsequent creativity. There was also an inverted-U shaped relationship between percentage of correct intervals and creativity: an intermediate proportion of correct intervals was identified as the "sweet spot" for creativity, while musical compositions with very few or too many correct intervals had the lowest creativity ratings. Finally, the P200 component in response to incorrect notes was predictive of creativity, suggesting a link between the neural correlates of learning, and creativity.

4.1. Behavioural and ERP indices of learning

Results showed that participants successfully learned the novel music grammar, as demonstrated by their post-test accuracy in recognizing notes belonging to the AMG. They also showed generalization of their knowledge to new melodies, suggesting internalization of governing rules of the novel grammar. Participants picked up the new grammar rules already from the pre-test (first exposure) session, as observed from their surprisal ratings. This is in line with previous studies showing that, irrespectively of the level of music training, humans are able to acquire statistical knowledge of novel music after a brief exposure, generalise to unheard melodies, and develop increased preferences for the learned grammar (Loui et al., 2009, 2010; Loui and Wessel, 2008). For example, a seminal study by Loui et al. (2010) used short melodies generated by an unfamiliar musical scale (Bohlen-Pierce scale) and found evidence for an acquired knowledge and an increased preference for this system as early as after 25 min of exposure. Further, Loui et al. (2009) provided neurophysiological evidence for acquired sensitivity to the statistical regularities of the novel music within the first 20 min of exposure, and generalization skills after 30 min of exposure. Finally, participants showed sensitivity to notes with different levels of probability, as evidenced by their surprisal judgements in the post-test which were higher for low-probability than high-probability notes, and even higher for incorrect notes, as those were more unpredictable given the AMG.

Increased sensitivity to differences in statistical properties of the notes was also reflected in the ERPs. In particular, the N100 amplitude was higher for incorrect than correct notes in both pre and post-test sessions, suggesting that learning started early on the first session. This early fronto-central negativity has been associated with violation of expectation in the context of melodic sequences (e.g., Carrión and Bly, 2008; Daikoku et al., 2015; Koelsch and Jentschke, 2010; Omigie et al., 2013; Zioga et al., 2016). In an EEG study demonstrating rapid probability learning of an unfamiliar musical system, Loui et al. (2009) reported enhanced early anterior negativity in response to deviant compared to standard chords. The authors suggested that this negativity might be an index of probability learning. The same effect has been observed for the P200 component, i.e. incorrect notes eliciting higher amplitudes compared to correct notes. This is in line with previous studies showing P200 increase in response to incongruent stimuli (e.g., Freunberger et al., 2007; Gruber and Müller, 2004). As the P200 has been associated with matching sensory inputs with stored memories (Freunberger et al., 2007), participants' enhanced P200 to incorrect notes in the post-test might suggest consolidated knowledge of the AMG. Finally, higher P3a in response to incorrect notes, especially in the post-test, can be also associated with the processing of invalid stimuli (e.g., Arthur and Starr, 1984; Knight et al., 1989; Loui et al., 2009; Yamaguchi and Knight, 1991). In a recent EEG study investigating expectation preference in jazz improvisers, Przysinda et al. (2017) reported higher P3a for low probability compared to high probability events. This has been interpreted as novelty detection and engagement with the unpredictable stimulus.

Table 1
Comparison between models. ANOVA comparing three regression models that predict creativity with those predictors: Model 1. Accuracy; 2. Accuracy and percentage of correct intervals (^2); 3. Accuracy, perc. corr. int. (^2), and P200 to incorrect notes.

Model/ Comparison	New predictor	Residual *df*	RSS	*F*	*p*
1/NA	Accuracy	31	9.805	NA	NA
2/2 vs. 1	Perc. corr. int. (^2)	29	7.484	5.829	.008
3/3 vs. 2	P200	28	5.573	9.602	.004
3/3 vs. 1	NA	28	5.573	7.087	.001

4.2. Neural mechanisms of encoding and retrieval of the learned music grammar

To investigate the neural mechanisms of encoding and retrieval of the newly learned music grammar, we analysed brain oscillations during the first exposure (pre-test) and the post-test, respectively. A cross-validation analysis showed that the delta power (2.5–4.5 Hz) in left temporal/central regions correlates with accuracy improvement, suggesting a potential neural mechanism of learning and acquisition of new knowledge. Additionally, high learners showed substantially higher delta power in the pre-test than low learners, but there was no group difference in the post-test. Trial level analysis showed that pre-test delta power was also enhanced for notes which were correctly judged in the post-test, providing corroborating evidence that it is an encoding signature. Phase analysis (see Supplementary materials) ruled out the possibility of delta oscillations reflecting a mere entrainment effect to the temporal characteristics of the stimuli, as intertrial phase coherence was significantly higher in the post-test.

Previous electrophysiological research has identified the role of delta, or so called slow-theta oscillations (around 2.5–5 Hz), in episodic memory processing in both rodents and humans (Lega et al., 2012; Watrous et al., 2011; Watrous et al., 2013). For example, in a verbal free recall task Lin et al. (2017) found that the slow-theta (2.5–5 Hz) power increased during successful item encoding. As the authors analysed intracranial EEG data, they were able to precisely locate this effect in the posterior

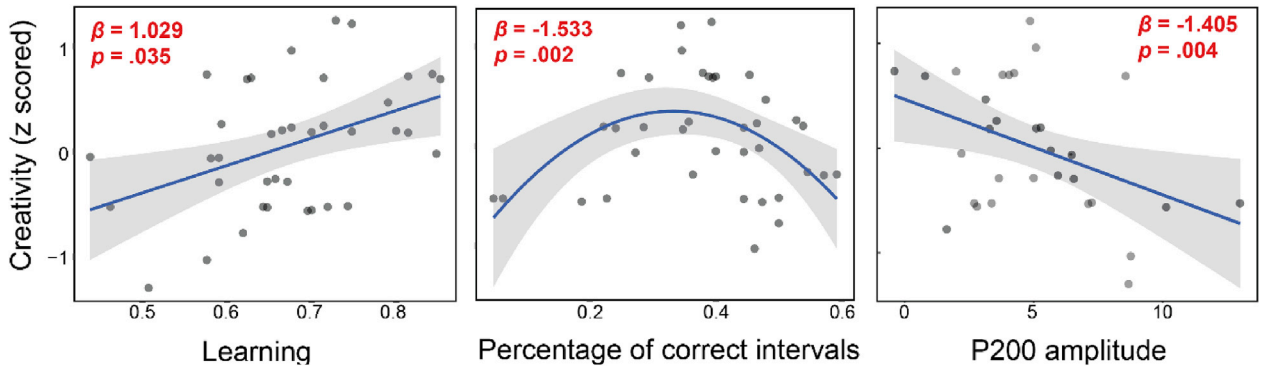


Fig. 10. The final model predicting perceived creativity (z-scored). Left: Linear term predicting creativity from learning (accuracy in the post-test). Middle: Quadratic term predicting creativity from the percentage of correct intervals of the musical compositions. Right: Linear term predicting creativity from the P200 amplitude in response to incorrect notes in the post-test.

hippocampus within 1 s after item presentation. The functional role of the 3-Hz slow-theta oscillation for memory formation has been contrasted to the 8-Hz fast-theta oscillation which has been more associated with locomotion (Seidenbecher et al., 2003). Lega et al. (2012) also demonstrated a distinct functionality between slow and fast-theta oscillations, where the former predict successful episodic memory encoding. We suggest that similar slow oscillatory activity is associated with encoding the statistical properties of melodies. In our study, the effects were widespread (left temporal, central and frontal regions) and as an EEG study we cannot know the exact source of our effects, which can be investigated in future studies.

In contrast to encoding, we analysed brain oscillations in the post-test to explore neural mechanisms of retrieval of the learned material. High-beta/low-gamma (18–32 Hz) band synchronization in left temporal areas and upper alpha (10–13 Hz) desynchronization in posterior areas correlated with post-test accuracy. High learners showed a higher beta synchronization and alpha desynchronization in response to correct notes during the post-test, compared to low learners. Considering that there was no group difference in the pre-test, we suggest that those effects reflect successful retrieval of the learned material. Previous research has primarily focused on the role of pre- and during-stimulus beta oscillations on subsequent memory success (Hanslmayr et al., 2008; Noh et al., 2014; Salari et al., 2012; Salari and Rose, 2016; Scholz et al., 2017). For example, Salari et al. (2012) found that stimuli presented in a state of increased beta power were more likely to be remembered later than those presented on decreased beta power, suggesting a role of the beta oscillations in memory encoding and formation. Interestingly, beta power increase has also been associated with inhibition of competing memories during selective retrieval (Jensen and Mazaheri, 2010; Klimesch et al., 2007; Waldhauser et al., 2012). In our case, there is a possibility that increased beta power in response to correct notes during the post-test reflects an active inhibition of competing representations in memory. As the melodies were suddenly interrupted without participants knowing when this would take place, it could be that they needed to inhibit the mental representation of the previous notes, in order to make an accurate judgement of the last note. Alternatively, participants needed to inhibit the pre-existing rules of the Western music which they have acquired through lifetime passive exposure, in order to be able to judge according to the rules of the AMG.

Alpha desynchronization has been typically associated with heightened attention as a mechanism to optimize perception of anticipated events, especially in visual tasks (e.g., Rohenkohl and Nobre, 2011; Sauseng et al., 2005; Mandikal Vasuki et al., 2017). For example, Rohenkohl and Nobre (2011) showed that alpha power fluctuated according to the tempo of rhythmic visual cues which induced temporal expectations about when a target event would occur. This was interpreted as a mechanism for biasing cortical excitability to enhance the perceptual processing of anticipated events. An alpha-band decrease was also reported in a visual task only when spatial attention was allocated, but not for unattended stimuli (Vázquez Marrufo et al., 2001). Additionally, there are studies showing that attention affects statistical learning (e.g., Baker et al., 2004; Jiménez and Mendez, 1999; Nissen and Bullemer, 1987; Turk-Browne et al., 2005). For example, a concurrent task during implicit learning of perceptual motor sequences and target-distractor pairings impaired performance on a serial reaction time task (Nissen and Bullemer, 1987). Importantly, Jiménez and Mendez (1999) demonstrated that selective attention to particular features but not others leads to implicit learning of only the attended features. Therefore, the reported larger alpha desynchronization in the high learners might reflect an enhanced attentional regulation during retrieval, suggesting potentially more efficient processing of the sensory input. We propose that future studies investigate alpha desynchronization by modulating the levels of attention during learning and retrieval of the learned material.

4.3. Human and computational assessments of musical compositions

Human evaluations of creativity were found to be positively correlated with judgements of both novelty and pleasantness. Novelty and adequacy have been considered the two main criteria for creativity (Sawyer, 2011). Pleasantness has been previously linked to creativity in various context (e.g., Amabile, 1982; Hickey, 2001). In her methodological study on the Consensual Assessment Technique (CAT) of creativity, Amabile (1982) demonstrated a very high correlation between subjective creativity and liking ratings ($r = 0.94$), suggesting that creativity judgements might be tightly coupled with assessments of aesthetic appeal. Positive affect has been also associated with creativity from the participant's point of view (Estrada et al., 1994). For example, participants in a positive emotional state have been found to produce significantly more correct solutions than individuals in a neutral and negative emotional state in the candle task (Duncker, 1945) and the Remote Associate Test, RAT (Mednick, 1968). In a study using humorous videotapes, positive affect has been associated with increased creativity ratings (Filipowicz, 2006). If positive mood states are reflected in the creative products, the latter might be more pleasant, and thus judged as more creative. Further, a study using the CAT technique found that children were unable to separate the concepts of liking and creativity, as musical compositions which were selected highest for liking were also judged as the most creative (Hickey, 2001). Our finding together with the above-mentioned evidence raise skepticism about how humans evaluate creativity (i.e. whether they are biased towards pleasant artefacts), and how independent these two measures are from each other.

As evidenced by the experts' ratings, judges appreciated a balance between correct and incorrect notes, where the highest creativity score was given to compositions with a moderate number of correct notes. This inverted U-shaped relationship indicates a "sweet spot" for creativity,

whereby an extreme adherence to stylistic rules is experienced as unexciting, while extreme novelty makes for incomprehensible music. A related inverted U-shaped curve between liking and complexity has previously been identified, where medium complexity stimuli were liked more than extremely low or highly complex stimuli, with a slight bias towards low complexity (Güçlütürk et al., 2016; for a review see Chmiel and Schubert, 2017). On surface, this inverted U association could be used to support Simonton's argument about expertise and creativity (1984). Simonton (1984) studied the relationship between formal education and creative productivity across life span, as measured by eminence in a specific field. Specifically, he analysed the space in reference works of 300 eminent individuals (e.g., Mozart, Einstein, Spinoza) in relation to their level of formal education. Simonton (1984) proposed an inverted U-shaped relationship between formal education and creative accomplishment. There is evidence that too much knowledge might hinder creativity (Frensch and Sternberg, 1989; Simonton, 2000): for example, Simonton (2000) analysed 911 operas by 59 classical composers and found that, in some instances, too much expertise was not beneficial for aesthetic success.

However, our results provide evidence for a positive linear association between expertise and creativity. We observed that the level of learning was found to predict subsequent creativity – specifically, high learners exhibited significantly higher creativity after learning compared to low learners and compared to the previous sessions. Furthermore, novelty ratings of high learners were stable across sessions, whereas novelty decreased for low learners. This could indicate that learning enables the generation of novel compositions, and not simple reproduction of the learned material. Studies of successful creative people have revealed that an average of 10 years of practice are required to achieve proficiency in a chosen field (Ericsson, 1996; Ericsson et al., 1993; Hayes, 1989). Bloom and Sosniak (1985) and Hayes (1989) suggested that a person must persevere with learning and practising a discipline for 10 years before they can make a breakthrough. This has been demonstrated in a broad range of domains, such as chess (Chase and Simon, 1973; Krogius, 1976), music (Sosniak, 1985), mathematics (Gustin, 1985), and sports (Bloom and Sosniak, 1985). Considering the aforementioned findings about the crucial role of deliberate practice for creativity, our findings suggest that learning is what enables creativity, through achieving higher novelty and an optimal, balanced amount of grammatical material. In contrary, weak learning might actually impair creativity by narrowing the person's range of ideas and thus reducing novelty. Framing this in a statistical context, the inverted U-shaped relationship between percentage of correct notes and creativity might suggest that a medium proportion of correct notes provides the appropriate balance between novelty and correctness. Our findings are in line with Boden's (2010, 2004) definition of the creative product as an artefact which is novel, surprising and adequate (correct). Overall, our findings offer a novel way to conceptualise and implement studies of creativity, by computing quantitative measures of grammaticality, in combination with human ratings of the various sub-dimensions of creativity, such as pleasantness, novelty and correctness.

We found that the P200 was enhanced in response to incorrect notes (compared to correct notes) after learning. Interestingly, the P200 in response to incorrect notes was found to be inversely correlated with creativity. These findings suggest that P200 might reflect increased sensitivity to the grammatical features of the grammar, associated with learning. Our findings could be explained in the context of preference/liking of unexpected events. In particular, Przysinda et al. (2017) found that jazz improvisers showed higher preference for unexpected chord progressions, compared to classical musicians and non-musicians who preferred expected chords. This is interpreted with Berlyne's (1971) theory that experts exhibit more complex structures compared to individuals with more domain-general knowledge. In our study, participants who achieved highly creative musical compositions might have a higher tolerance or even preference for incorrect notes, compared to less creative people who react strongly to incorrect notes. Our findings expand on Przysinda et al. (2017) as they show that more creative people might have tolerance not only for unexpected notes, but also for incorrect, wrong notes. Future studies are recommended to investigate whether people who are more sensitive to mistakes are less creative.

On the other hand, the oscillatory correlates of both encoding and retrieval were not predictive of creativity. This finding suggests that creativity may not critically depend on neural processes directly linked to encoding and retrieving, but might be associated with more complex processes (e.g., coupling between oscillations, connectivity patterns). One possibility is that the processes involved in encoding and retrieving learned information are not directly associated with the capacity of generating novelty based on this knowledge. Learning was found to be crucial for creativity, however, other cognitive and affective processes which took place during the training, the test sessions, or the preparation of the musical compositions might have played a critical role for creativity as well. Further, there is a possibility that the neural measures derived from the EEG recordings are not the appropriate ones that could index how learning relates to creativity. Different EEG experiments might need to be designed to test questions about encoding/retrieval of learned material vs. questions about creativity. Future studies on learning of an AMG and subsequent creativity with professional musicians would be appropriate to investigate how long-term musical knowledge might be in competition with a newly learned musical style.

This research is not without limitations. First, as much as we tried to create an ecologically valid paradigm, musical training takes much more time in the context of real life. Second, individual differences in terms of learning were not taken into account, i.e. we assumed that participants learned during training in a homogeneous manner across sessions, but it could be that some sequences were not learned the first time they were presented. Further, it might be that differences in the experimental design between the intermediate tests and the pre- and post-tests influenced the results, therefore the experimental design could be improved to ensure identical test procedures throughout the experiment. Finally, non-invasive brain stimulation experiments might be useful to provide some causal evidence for low frequency oscillations as a neural mechanism of encoding, by manipulating learning through stimulation at that particular frequency.

5. Conclusions

Our study introduced a novel experimental paradigm combining behavioural, electrophysiological and computational evidence to investigate the relationship between novel music learning and creativity. We further attempted to simulate learning from scratch over multiple sessions. Instead of finding only the effects of learning on the brain, we identified the neural correlates of learning during encoding of the learned material, i.e. the mechanisms which were responsible for the learning. Finally, we revealed direct associations between behavioural and neural measures, and human judgements of creativity, offering novel contributions to the investigation of creativity.

Acknowledgements

We thank Khadija Rita Khatun for help with EEG data collection. Ioanna Zioga is supported by a doctoral studentship from the Department of Biological and Experimental Psychology at the School of Biological and Chemical Sciences, Queen Mary University of London. Peter Harrison is supported by a doctoral studentship from the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1). We also thank the anonymous reviewers for their constructive comments towards the improvement of the paper. The authors declare no competing financial interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116311.

References

Abla, D., Katahira, K., Okanoya, K., 2008. On-line assessment of statistical learning by event-related potentials. J. Cogn. Neurosci. 20 (6), 952–964.
Abla, D., Okanoya, K., 2009. Visual statistical learning of shape sequences: an ERP study. Neurosci. Res. 64 (2), 185–190.
Amabile, T., 1982. Social psychology of creativity: a consensual assessment technique. J. Personal. Soc. Psychol. 43 (5), 997–1013.
Arthur, D., Starr, A., 1984. Task-relevant late positive component of the auditory event-related potential in monkeys resembles P300 in humans. Science 223 (4632), 186–188.
Baker, C.I., Olson, C.R., Behrmann, M., 2004. Role of attention and perceptual grouping in visual statistical learning. Psychol. Sci. 15 (7), 460–466. https://doi.org/10.1111/j.0956-7976.2004.00702.x.
Beaty, R., 2015. The neuroscience of musical improvisation. Neurosci. Biobehav. Rev.
Begleiter, R., El-Yaniv, R., Yona, G., 2004. On prediction using variable order Markov models. J. Artif. Intell. Res. 22, 385–421.
Bengtsson, S.L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., Ullén, F., 2005. Extensive piano practicing has regionally specific effects on white matter development. Nat. Neurosci. 8 (9), 1148–1150. https://doi.org/10.1038/nn1516.
Berlyne, D., 1971. Aesthetics and Psychobiology. Appleton-Century-Crofts, New York.
Bloom, B.S., Sosniak, L.A., 1985. Developing Talent in Young People. Ballantine Books, New York, NY.
Boden, M., 2010. Creativity and Art: three Roads to Surprise. Oxford University Press, Oxford.
Brainard, D., Vision, S., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436.
Bunton, S., 1997. Semantically motivated improvements for PPM variants. Comput. J. 40 (2_and_3), 76–93.
Carrión, R., Bly, B., 2008. The effects of learning on event-related potential correlates of musical expectancy. Psychophysiology 45 (5), 759–775. https://doi.org/10.1111/j.1469-8986.2008.00687.x.
Chase, W., Simon, H., 1973. Perception in chess. Cogn. Psychol. 4 (1), 55–81.
Chen, S., Goodman, J., 1999. An empirical study of smoothing techniques for language modeling. Comput. Speech Lang 13 (4), 359–394.
Chmiel, A., Schubert, E., 2017. Back to the inverted-U for music preference: a review of the literature. Psychol. Music 45 (6), 886–909. https://doi.org/10.1177/0305735617697507.
Christiansen, M.H., 2019. Implicit statistical learning: a tale of two literatures. Top. Cogn. Sci. 11 (3), 468–481. https://doi.org/10.1111/tops.12332.
Daikoku, T., Yatomi, Y., Yumoto, M., 2015. Statistical learning of music-and language-like sequences and tolerance for spectral shifts. Neurobiol. Learn. Mem. 118, 8–19.
Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21.
Dienes, Z., Longuet-Higgins, C., 2004. Can musical transformations be implicitly learned? Cogn. Sci. 28 (4), 531–558.
Duncker, K., 1945. On problem solving. Psychol. Monogr. 58 (5). Whole No. 270.
Egermann, H., Pearce, M.T., Wiggins, G.A., McAdams, S., 2013. Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. Cognit. Affect Behav. Neurosci. 13 (3), 533–553.
Ericsson, K.A. (Ed.), 1996. The Road to Expert Performance: Empirical Evidence from the Arts and Sciences, Sports, and Games. Erlbaum, Mahwah, NJ.
Ericsson, K.A., Krampe, R.T., Tesch-Romer, C., 1993. The role of deliberate practice in the acquisition of expert performance. Psychol. Rev. 100 (3), 363–406.
Estrada, C.A., Isen, A.M., Young, M.J., 1994. Positive affect improves creative problem solving and influences reported source of practice satisfaction in physicians. Motiv. Emot. 18 (4), 285–299. https://doi.org/10.1007/BF02856470.
Filipowicz, A., 2006. From positive affect to creativity: the surprising role of surprise. Creativ. Res. J. 18 (2), 141–152. https://doi.org/10.1207/s15326934crj1802_2.
Frensch, P.A., Sternberg, R.J., 1989. Expertise and intelligent thinking: when is it worse to know better. Adv. Psychol. Hum. Intell. 5, 157–188.
Freunberger, R., Klimesch, W., Doppelmayr, M., Höller, Y., 2007. Visual P2 component is related to theta phase-locking. Neurosci. Lett. 426 (3), 181–186.
Gruber, T., Müller, M., 2004. Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. Cerebr. Cortex 15 (1), 109–116.
Güçlütürk, Y., Jacobs, R.H.A.H., Lier, R.van, 2016. Liking versus complexity: decomposing the inverted U-curve. Front. Hum. Neurosci. 10, 112. https://doi.org/10.3389/fnhum.2016.00112.
Gustin, W.C., 1985. The development of exceptional research mathematicians. In: Developing Talent in Young People. Ballantine Books, New York, NY, pp. 139–192.
Hanslmayr, S., Spitzer, B., Bäuml, K., 2008. Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. Cerebr. Cortex 19 (7), 1631–1640.
Hayes, J.R., 1989. Cognitive processes in creativity. In: Glover, J.A., Ronning, R.R., Reynolds, C.R. (Eds.), Handbook of Creativity. Plenum Press, New York, NY, pp. 135–145.
Hickey, M., 2001. An application of Amabile's Consensual Assessment Technique for rating the creativity of children's musical compositions. J. Res. Music Educ. 49 (3), 234–244. https://doi.org/10.2307/3345709.
Huron, D.B., 2006. Sweet Anticipation: Music and the Psychology of Expectation. MIT Press.
Jensen, O., Mazaheri, A., 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front. Hum. Neurosci. 4 https://doi.org/10.3389/fnhum.2010.00186.
Jiménez, L., Mendez, C., 1999. Which attention is needed for implicit sequence learning? J. Exp. Psychol. Learn. Mem. Cogn. 25 (1), 236.
Juslin, P.N., 2019. Musical Emotions Explained. Oxford University Press, Oxford.
Klimesch, W., Sauseng, P., Hanslmayr, S., 2007. EEG alpha oscillations: the inhibition–timing hypothesis. Brain Res. Rev. 53 (1), 63–88.
Knight, R.T., Scabini, D., Woods, D.L., Clayworth, C.C., 1989. Contributions of temporal-parietal junction to the human auditory P3. Brain Res. 502 (1), 109–116. https://doi.org/10.1016/0006-8993(89)90466-6.
Koelsch, S., Busch, T., Jentschke, S., Rohrmeier, M., 2016. Under the hood of statistical learning: a statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. Sci. Rep. 6 (1), 19741. https://doi.org/10.1038/srep19741.
Koelsch, S., Jentschke, S., 2010. Differences in electric brain responses to melodies and chords. J. Cogn. Neurosci. 22 (10), 2251–2262.
Krogius, N., 1976. Psychology in Chess. RHM Press, New York, NY.
Lega, B.C., Jacobs, J., Kahana, M., 2012. Human hippocampal theta oscillations and the formation of episodic memories. Hippocampus 22 (4), 748–761. https://doi.org/10.1002/hipo.20937.
Lieberman, M.D., Chang, G.Y., Chiao, J., Bookheimer, S.Y., Knowlton, B.J., 2004. An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. J. Cogn. Neurosci. 16 (3), 427–438.
Limb, C., Braun, A., 2008. Neural substrates of spontaneous musical performance: an FMRI study of jazz improvisation. PLoS One 3 (2), e1679.
Lin, J.-J., Rugg, M.D., Das, S., Stein, J., Rizzuto, D.S., Kahana, M.J., Lega, B.C., 2017. Theta band power increases in the posterior hippocampus predict successful episodic memory encoding in humans. Hippocampus 27 (10), 1040–1053. https://doi.org/10.1002/hipo.22751.
Liu, S., Chow, H., Xu, Y., Erkkinen, M., Swett, K., 2012. Neural correlates of lyrical improvisation: an fMRI study of freestyle rap. Sci. Rep. 2, 834.
Lopata, J.A., Nowicki, E.A., Joanisse, M.F., 2017. Creativity as a distinct trainable mental state: an EEG study of musical improvisation. Neuropsychologia 99, 246–258. https://doi.org/10.1016/j.neuropsychologia.2017.03.020.
Loui, P., 2012. Learning and liking of melody and harmony: further studies in artificial grammar learning. Top. Cogn. Sci. 4 (4), 554–567.
Loui, P., Wessel, D.L., 2008. Learning and liking an artificial musical system: effects of set size and repeated exposure. Music. Sci.: J. Eur. Soc. Cognitive Sci. Music 12 (2), 207.
Loui, P., Wessel, D.L., Kam, C.L.H., 2010. Humans rapidly learn grammatical structure in a new musical scale. Music Percept. Interdiscip. J. 27 (5), 377–388.
Loui, P., Wu, E.H., Wessel, D.L., Knight, R.T., 2009. A generalized mechanism for perception of pitch patterns. J. Neurosci. 29 (2), 454–459.
Luft, C.D.B., Baker, R., Goldstone, A., Zhang, Y., Kourtzi, Z., 2016. Learning temporal statistics for sensory predictions in aging. J. Cogn. Neurosci. 28 (3), 418–432.
Mandikal Vasuki, P.R., Sharma, M., Ibrahim, R.K., Arciuli, J., 2017. Musicians' online performance during auditory and visual statistical learning tasks. Front. Hum. Neurosci. 11 https://doi.org/10.3389/fnhum.2017.00114.
Mednick, S.A., 1968. The Remote associates test. J. Creat. Behav. 2 (3), 213–214. https://doi.org/10.1002/j.2162-6057.1968.tb00104.x.
Misyak, J.B., Christiansen, M.H., Tomblin, J.B., 2010. Sequential expectations: the role of prediction-based learning in language. Top. Cogn. Sci. 2 (1), 138–153.
Müllensiefen, D., Gingras, B., Musil, J., Stewart, L., 2014. The musicality of non-musicians: an index for assessing musical sophistication in the general population. PLoS One 9 (2), e89642.
Narmour, E., 1992. The Analysis and Cognition of Melodic Complexity: the Implication-Realization Model. University of Chicago Press.
Nissen, M., Bullemer, P., 1987. Attentional requirements of learning: evidence from performance measures. Cogn. Psychol. 19 (1), 1–32.
Noh, E., Herzmann, G., Curran, T., de Sa, V., 2014. Using single-trial EEG to predict and analyze subsequent memory. Neuroimage 84, 712–723.
Omigie, D., Pearce, M.T., Williamson, V.J., Stewart, L., 2013. Electrophysiological correlates of melodic processing in congenital amusia. Neuropsychologia 51 (9), 1749–1762. https://doi.org/10.1016/j.neuropsychologia.2013.05.010.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 1. *2011*.
Pachet, F., 1999. Surprising harmonies. Int. J. Comput. Anticipatory Syst. (4), 139–161.
Pearce, M.T., 2005. The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition. Doctoral dissertation, City University London.
Pearce, M.T., 2018. Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. Ann. N. Y. Acad. Sci. 1423 (1), 378–395. https://doi.org/10.1111/nyas.13654.
Pearce, M.T., Conklin, D., Wiggins, G.A., 2005. Methods for combining statistical models of music. In: Wiil K., U. (Ed.), Computer Music Modelling and Retrieval. Springer Verlag, Heidelberg, Germany, pp. 295–312.
Pearce, M.T., Müllensiefen, D., Wiggins, G.A., 2010a. The role of expectation and probabilistic learning in auditory boundary perception: a model comparison. Perception 39 (10), 1365–1389.
Pearce, M.T., Ruiz, M.H., Kapasi, S., Wiggins, G.A., Bhattacharya, J., 2010b. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. Neuroimage 50 (1), 302–313.
Pearce, M.T., Wiggins, G.A., 2007. Evaluating cognitive models of musical composition. In: Proceedings of the 4th International Joint Workshop on Computational Creativity. University of London, Goldsmiths, pp. 73–80.
Pinho, A., Manzano, Ö.de, Fransson, P., Eriksson, H., Ullén, F., 2014. Connecting to create: expertise in musical improvisation is associated with increased functional connectivity between premotor and prefrontal areas. J. Neurosci. 34 (18), 6156–6163.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118 (10), 2128–2148. https://doi.org/10.1016/J.CLINPH.2007.04.019.

Pothos, E.M., 2007. Theories of artificial grammar learning. Psychol. Bull. 133 (2), 227–244.

Przysinda, E., Zeng, T., Maves, K., Arkin, C., Loui, P., 2017. Jazz musicians reveal role of expectancy in human creativity. Brain Cogn. 119, 45–53.

Reber, A.S., 1993. Implicit Learning and Knowledge: an Essay on the Cognitive Unconscious. Oxford University Press, New York, NY.

Rohenkohl, G., Nobre, A.C., 2011. Alpha oscillations related to anticipatory attention follow temporal expectations. J. Neurosci. 31 (40), 14076–14084. https://doi.org/10.1523/JNEUROSCI.3387-11.2011.

Rohrmeier, M.A., Cross, I., 2014. Modelling unsupervised online-learning of artificial grammars: linking implicit and statistical learning. Conscious. Cognit. 27, 155–167.

Rohrmeier, M.A., Koelsch, S., 2012. Predictive information processing in music cognition. A critical review. Int. J. Psychophysiol. 83 (2), 164–175.

Rohrmeier, M.A., Rebuschat, P., 2012. Implicit learning and acquisition of music. Top. Cogn. Sci. 4, 525–553. https://doi.org/10.1111/j.1756-8765.2012.01223.x.

Rohrmeier, M.A., Rebuschat, P., Cross, I., 2011. Incidental and online learning of melodic structure. Conscious. Cognit. 20 (2), 214–222.

Runco, M., Jaeger, G., 2012. The standard definition of creativity. Creativ. Res. J. 24 (1), 92–96. https://doi.org/10.1080/10400419.2012.650092.

Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. Science 274 (5294), 1926–1928.

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, 27–52.

Saffran, J.R., Reeck, K., Niebuhr, A., Wilson, D., 2005. Changing the tune: the structure of the input affects infants' use of absolute and relative pitch. In: Developmental Science, vol. 8.

Salari, N., Büchel, C., Rose, M., 2012. Functional dissociation of ongoing oscillatory brain states. PLoS One 7 (5), e38090. https://doi.org/10.1371/journal.pone.0038090.

Salari, N., Rose, M., 2016. Dissociation of the functional relevance of different pre-stimulus oscillatory activity for memory formation. Neuroimage 125, 1013–1021.

Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., et al., 2005. A shift of visual spatial attention is selectively associated with human EEG alpha activity. Eur. J. Neurosci. 22 (11), 2917–2926. https://doi.org/10.1111/j.1460-9568.2005.04482.x.

Sawyer, R.K., 2011. Explaining Creativity: the Science of Human Innovation, second ed. Oxford university press, Inc, New York.

Scholz, S., Schneider, S.L., Rose, M., 2017. Differential effects of ongoing EEG beta and theta power on memory formation. PLoS One 12 (2), e0171913. https://doi.org/10.1371/journal.pone.0171913.

Seidenbecher, T., Laxmi, T., Stork, O., Pape, H., 2003. Amygdalar and hippocampal theta rhythm synchronization during fear memory retrieval. Science 301 (5634), 846–850.

Simonton, D.K., 1984. Genius, Creativity and Leadership. Harvard University Press, Cambridge, MA.

Simonton, D.K., 2000. Creative development as acquired expertise: theoretical issues and an empirical test. Dev. Rev. 20 (2), 283–318.

Simonton, D.K., 2012. Taking the U.S. Patent Office criteria seriously: a quantitative three-criterion creativity definition and its implications. Creativ. Res. J. 24 (2–3), 97–106. https://doi.org/10.1080/10400419.2012.676974.

Sosniak, L.A., 1985. Learning to be a concert pianist. Dev. Talent Young People 1, 19–67.

Tillmann, B., McAdams, S., 2004. Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis) similarities. J. Exp. Psychol. Learn. Mem. Cogn. 30 (5), 1131–1142.

Turk-Browne, N., Jungé, J., Scholl, B., 2005. The automaticity of visual statistical learning. J. Exp. Psychol. Gen. 134 (4), 552.

Varshney, L., Pinel, F., Varshney, K., Schörgendorfer, A., Chee, Y., 2013. Cognition as a part of computational creativity. In: 2013 IEEE 12th International Conference on Cognitive Informatics and Cognitive Computing. IEEE, pp. 36–43.

Vázquez Marrufo, M., Vaquero, E., Cardoso, M.J., Gómez, C.M., 2001. Temporal evolution of α and β bands during visual spatial attention. Cogn. Brain Res. 12 (2), 315–320. https://doi.org/10.1016/S0926-6410(01)00025-8.

Waldhauser, G., Johansson, M., Hanslmayr, S., 2012. Alpha/beta oscillations indicate inhibition of interfering visual memories. J. Neurosci. 32 (6), 1953–1961.

Watrous, A.J., Fried, I., Ekstrom, A.D., 2011. Behavioral correlates of human hippocampal delta and theta oscillations during navigation. J. Neurophysiol. 105 (4), 1747–1755. https://doi.org/10.1152/jn.00921.2010.

Watrous, A.J., Lee, D.J., Izadi, A., Gurkoff, G.G., Shahlaie, K., Ekstrom, A.D., 2013. A comparative study of human and rat hippocampal low-frequency oscillations during spatial navigation. Hippocampus 23 (8), 656–661. https://doi.org/10.1002/hipo.22124.

Yamaguchi, S., Knight, R.T., 1991. P300 generation by novel somatosensory stimuli. Electroencephalogr. Clin. Neurophysiol. 78 (1), 50–55. https://doi.org/10.1016/0013-4694(91)90018-Y.

Zioga, I., Luft, C.D.B., Bhattacharya, J., 2016. Musical training shapes neural responses to melodic and prosodic expectation. Brain Res. 1650, 267–282.