Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

__________________________________   ________________

Rohini Murugan   Date

Assessing the Nature of Knowledge that is acquired in an Artificial Grammar Learning Paradigm

By

Rohini Murugan

Master of Arts

Psychology

Benjamin Wilson, PhD

Advisor

Harold Gouzoules, PhD

Committee Member

Robert Hampton, PhD

Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D, MPH

Dean of the James T. Laney School of Graduate Studies

Date

Assessing the Nature of Knowledge that is acquired in an Artificial Grammar Learning Paradigm

By

Rohini Murugan
B.Sc., Women’s Christian College, 2017
MS., Indian Institute of Science, 2021

Advisor: Benjamin Wilson, PhD

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfilment of the requirements for the degree of
Master of Arts
in Psychology
2023

Abstract

Assessing the Nature of Knowledge that is acquired in an Artificial Grammar Learning Paradigm

By Rohini Murugan

Languages follow grammatical rules, which determine the order of words in sentences. One of the ways in which infants learn language is by learning these rules implicitly from the speech that they are exposed to. This process is called implicit learning and has been studied using artificial grammar learning paradigms. Typically, in these experiments, learning is assumed to be implicit. However, much of our formal education involves a different process called explicit learning. Though implicit and explicit learning has been studied in humans, the measures used to identify these mechanisms have certain limitations. Further, very little is known about how nonhuman animals learn similar rules in an AGL task and these measures are difficult to be adapted to nonhuman animal tasks. Thus, we combined an AGL task with a metacognitive measure that has been previously used in nonhuman animals to assess the nature of knowledge in humans. Participants were initially exposed to sequences of visual symbols generated by an artificial grammar, followed by a testing phase in which they classified sequences as either grammatical or ungrammatical. On a subset of trials (choice trials), participants were given the option to either take or skip this classification. If participants had explicit knowledge of the grammar, they should be more likely to take trials when they are confident and skip when they are not. Participants learned the grammar and reported explicit knowledge of the grammar. However, they performed similarly in forced and choice trials, suggesting that this metacognitive measure does not measure the explicitness of their knowledge of this grammar.

Assessing the Nature of Knowledge that is acquired in an Artificial Grammar Learning Paradigm

By

Rohini Murugan

B.Sc., Women's Christian College, 2017

MS., Indian Institute of Science, 2021

Advisor: Benjamin Wilson, PhD

A thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfilment of the requirements for the degree of

Master of Arts

in Psychology

2023

Contents

Tables

Introduction

Our environment is not random. It is replete with events that are predictable. For example, lightning is always followed by thunder. It is helpful for any organism to be sensitive to this predictability. Beyond predictable patterns in the environment, language represents another system which contains many statistical regularities.

All languages across different cultures show statistically predictable regularities. Languages comprise smaller units like syllables and phonemes that combine recursively into larger units like words, phrases, and sentences. This recursive process gives rise to certain regularities in language that are statistically predictable. An example of such regularity in the English language is the probability of co-occurrence of syllables in words. Most of us are exposed to the word '*science*' more frequently than '*scion*'. Because of this unequal exposure, when we come across the syllable '*sci*', we would predict it to be mostly followed by the syllable '*ence*' than by '*ion*'. This regularity is quantified by calculating the probability of the occurrence of an event B ('*ence'*) given the occurrence of event A ('*sci'*) and is called the transitional probability (TP). Transition probabilities between syllables across different words have been shown to help in learning the boundaries between words and thus identifying words within a rapid speech stream (Pelucchi et al., 2009; Perruchet, 2019; Saffran et al., 1996). For example, in the phrase, '*pretty baby*', the probability of transition between the syllables '*pre*' and '*tty*' is higher than the probability of transition between the syllables '*tty*' and '*ba*'. This difference in TPs between the syllable pairs is one of the ways to learn the words '*pretty*' and '*baby*', and the boundary between them. This ability to learn statistical patterns is termed 'statistical learning' and has been observed experimentally in humans (Conway & Christiansen, 2005, 2009; Pelucchi et al., 2009; Saffran et al., 2008; Saffran et al., 1996; Turesson & Ghazanfar, 2011; Turk-Browne et al., 2009) and in nonhuman animals (Fitch & Hauser, 2004; Milne et al., 2018; Newport et al., 2004; Ravignani et al., 2013; Saffran et al., 2008; Sonnweber et al., 2015; Wilson et al., 2015) using Artificial Grammar Learning (AGL) paradigms. These paradigms have been used to study the type of statistical patterns learnt during the task but not how they are learnt. These studies typically assume that learning occurs without conscious awareness or explicit knowledge of the rules (Christiansen, 2019; Pelucchi et al., 2009; Perruchet, 2019; Perruchet & Pacton, 2006; Saffran et al., 1996) and it has even been suggested that statistical learning and implicit learning are two approaches that study the same underlying mechanisms (Perruchet & Pacton, 2006).

Implicit and Explicit Learning

However, learning is not always implicit. Much of our formal education involves a different process called explicit learning. Clinical studies suggest that implicit and explicit learning mechanisms are two separate mechanisms that could be recruited in an AGL task. Evidence from these studies demonstrates that amnesic patients with damage to their Medial Temporal Lobe (MTL) and hippocampus and thus lost their ability to store and use explicit knowledge perform as well as the normal population in the AGL tasks (Knowlton et al., 1992; Knowlton & Squire, 1994). Since the patients cannot encode new explicit knowledge, the only way they could have performed well in the task would be by developing new implicit knowledge, which could have been the product of implicit learning. Further, they also performed worse than the normal population in AGL task conditions that required the recruitment of explicit mechanisms like recognition (Knowlton et al., 1992) suggesting that patients were unable to use explicit mechanisms that were otherwise available to the normal population to perform in an AGL task. These studies demonstrate that both implicit and explicit learning systems could be recruited by participants in AGL tasks as opposed to only implicit learning as is commonly assumed. This raises the critical need of measures that could identify these learning mechanisms.

Artificial Grammar Learning Paradigms

Much of the behavioural work on identifying implicit and explicit learning systems in the context of language comes from artificial grammar learning (AGL) paradigms where the type of learning is inferred by assessing whether the knowledge is implicit or explicit (de Vries et al., 2010; Dienes et al., 1991; Gomez & Gerken, 1999; Knowlton et al., 1992; Knowlton & Squire, 1994; Reber, 1967; Saffran & Kirkham, 2018). These paradigms use artificial grammars which are sets of rules that govern the ordering of arbitrary elements into sequences. A typical AGL task has an exposure phase and a grammaticality judgement (GJ) phase. In the exposure phase, participants are exposed to the grammar by being presented with grammatical sequences. Following the exposure phase, they are told about the presence of rules in the sequences they had just observed. They are then presented with both grammatical and ungrammatical sequences and are instructed to judge whether the presented sequence adhered to the same rules as the sequences they had observed in the exposure phase. This phase is called the grammaticality judgement phase. Following this phase, the participants are typically asked to report the structure of the grammar verbally or are asked questions regarding the structure of the grammar. Performance above chance in the grammaticality judgement phase indicates that the participants had learned the grammar, and the learning mechanism that was recruited to learn the grammar is usually inferred by assessing the nature of their knowledge evidenced by the verbal reports. If the participants show learning but cannot report the grammar rules in the verbal reports, then they are suggested to have acquired implicit knowledge. On the other hand, if the participants show learning in the grammaticality judgement phase but could also report the structure of the grammar in the verbal reports, then they are said to have acquired explicit knowledge (Batterink et al., 2015; Reber, 1967, 1976, 1989; Reber et al., 1991; Rebuschat & Williams, 2012).

Another measure that has been used in the AGL literature to assess the nature of knowledge involves asking participants to make confidence ratings about their decisions (Channon et al., 2002; Dienes et al., 1995a; Dienes & Berry, 1997; Dienes & Scott, 2005; Rebuschat & Williams, 2012; Tunney & Shanks, 2003). Typically, in these studies, the participants are instructed to rate their confidence after every trial in the grammaticality judgement task. Explicit knowledge is inferred based on the correlation between the participant's accuracy and their confidence; if they have explicit knowledge of the grammar, then they would be more confident in trials in which they are correct and less confident in trials in which they are incorrect, resulting in a high performance-confidence correlation. On the other hand, if they have implicit knowledge of the grammar, then they would show no correlation between their performance and confidence. However, despite their importance in identifying the learning mechanisms, these measures have certain limitations.

Limitations of methods used in Artificial Grammar Learning Paradigms

With respect to verbal reports, in addition to being subjective measures, they are also unreliable because, the absence of evidence of explicit knowledge in verbal reports does not necessarily mean that participants did not have explicit knowledge of the grammar. They could have found it hard to articulate the rules of the grammar or could have been unmotivated to report it (Dienes & Perner, 1999). Most studies (Reber, 1976; Yang & Li, 2012) assume that if participants are not informed about the rules before the task, then they would learn the grammar implicitly. No independent measures were used to test this assumption in these studies. It is plausible that, even if not informed about the rules by the experimenter, the participants could have explicitly looked for rules in the sequences. This would have resulted in the acquisition of explicit knowledge in contrast to the tacitly assumed implicit knowledge as assumed by most AGL studies. Verbal reports also could not be used in nonhuman animal research since nonhuman animals cannot verbalize. Though confidence ratings is a reliable measure to assess the explicitness of knowledge, it is difficult to train monkeys to provide a measure of their confidence with the same level of granularity that typical AGL tasks in humans do (rated on a scale of 1-100). Thus, an essential step for effective cross-species comparisons would be a novel measure that is reliable and that could be used in both humans and nonhuman animals to assess similarities and differences in how humans and nonhuman animals use implicit and explicit learning mechanisms.

Metacognition as a measure to assess the explicitness of knowledge

Methods that are used to assess metacognition are one of the ways to assess implicit and explicit knowledge. Metaknowledge is the secondary level of representation of an existing knowledge state. The underlying cognitive process involved in accessing that metaknowledge is termed metacognition. In an AGL task, participants classifying sequences above chance demonstrate that they have acquired knowledge of the grammar. If they had acquired an implicit knowledge of the grammar, then they would not be aware that they know the grammar. Thus, they would not have a secondary representation or metaknowledge of their knowledge of the grammar. On the other hand, if they possess explicit knowledge of the grammar, then they would be aware of their knowledge state, resulting in them having metaknowledge of their knowledge of the grammar. Implicit knowledge could not be accessed by metacognition, while explicit knowledge could be accessed. Thus, testing whether participants have metaknowledge or not is one of the ways to assess the implicitness of their knowledge. Measures like verbal reports and performance-confidence correlation that were discussed above could be considered as measures that test this metaknowledge.

One experimental paradigm that has been developed and widely used to study metacognition in nonhuman animals is a paradigm that implemented the option to ‘skip’ a trial if the participant was uncertain about their response. This paradigm (which we mention hereafter as the ‘forced/choice’ paradigm) has been used effectively across different species like rhesus macaques, pigeons, rats, and dolphins to test whether they can monitor their knowledge states (Adams & Santi, 2011; Foote & Crystal, 2007; Hampton, 2001; Inman & Shettleworth, 1999; Shields et al., 1997; Smith et al., 1995). In one such study, Hampton (2001) investigated metacognition in rhesus macaques using this paradigm. The monkeys were trained to respond to a delayed match to sample task. In this task, they were shown a stimulus, and after a delay they must correctly identify the stimulus among three distractors. The task becomes more difficult as the delay becomes longer. In two-thirds of the trials, after seeing the stimulus, they are given the option to either skip or take the trial (choice trials). The option of skipping the trial would result in a guaranteed, low-value reward. However, if they chose to take the trial and got it correct, they received a higher-value reward, while if they made an incorrect response, they received a timeout and no reward. The monkeys showed higher performance in choice trials as compared to forced trials, because they were able to preferentially skip hard trials while successfully completing easy ones. This difference in accuracy was observed in the monkeys demonstrated the presence of metacognitive ability. Though this paradigm has been used to successfully test cognitive processes like memory in nonhuman animals, they have not yet been employed in AGL paradigms to study statistical learning, both in humans and nonhuman animals.

Thus, in a series of two experiments, we combined the forced/choice paradigm with a traditional AGL task to develop an experimental paradigm that could test the extent to which humans rely on implicit and explicit knowledge during artificial grammar learning. In this new experimental setup, we introduced a new trial condition, forced and choice trials in the grammaticality judgement phase of a typical AGL task. We hypothesized that if participants acquire explicit knowledge of the grammar, then they will perform better in the choice trials than in the forced trials. On the other hand, if the participants acquire implicit knowledge of the grammar, they will perform similarly in both, forced and choice trials. We also tested and validated this forced/choice accuracy difference as a measure of metaknowledge by comparing it against the historically used metacognitive measures like confidence ratings and verbal reports. This measure could provide a novel independent way to assess knowledge in AGL tasks. Further, since this measure has also been used in nonhuman animals, it could open interesting venues for future comparative research to identify similarities and differences in the learning mechanisms used by humans and nonhuman animals to learn patterns.

Experiment 1

Participants

We recruited 41 participants (Mean age: 19.12 years; 14 males and 27 females) for this experiment. All participants were Emory University undergraduate students and received one research credit for their participation. The experiment was approved by Emory University IRB (ID: STUDY00003577).

Stimuli

Artificial Grammar

Artificial grammars are sets of rules that dictate the ordering of elements in a sequence. They can be designed to capture specific properties of natural language. The artificial grammar we used here generates sequences that mimic certain aspects of language (Figure 1A). Firstly, in language, pairs of syllables that occur together within words typically co-occur at higher frequencies than pairs of syllables that occur across words. For example, in the phrase '*Happy birthday'* the syllable '*Hap*' precedes the syllable '*py'* with a higher frequency than '*py'* precedes '*birth'*. This can be quantified by calculating Transitional Probabilities (TPs) for each such transition. TPs are calculated by the formula

$$P(B|A) = \frac{Probability\ of\ A\ and\ B\ occuring\ together}{Probability\ of\ A}$$

which gives P(B|A), the probability of transition from A to B. In the previous example, since the probability of co-occurrence of syllables, *'Hap'* and *'py'* is higher than that of the syllables *'py'* and *'birth'*, the transitional probability of the first pair of syllables, *'Hap-py'* would be higher than the second pair of syllables, *'py-birth'*. Indeed, the relatively higher TPs within words than between words is one of the critical properties of language that helps infants initially identify the boundaries between two words, and thus words themselves (Saffran et al., 1996).

Our grammar generates sequences that contain both these kinds of syllable pairs. Some stimuli pairs always occur together in a fixed order and therefore have a TP equal to 1.0. These stimuli pairs mimic the syllable pairs within words like *'Hap-py'* and we refer to them as 'within-chunk transitions'. Some stimuli pairs mimic syllables between two words. In the previous example

‘*Happy*’ may be more likely to be followed by ‘*birthday’* than *‘baby’*, despite both ‘*Happy birthday’* and ‘*Happy baby*’ being perfectly grammatical phrases. Therefore, our grammar generates transitions with TPs lower than the ‘within-chunk transitions’ (0.67 and 0.33), and we refer to these stimuli as ‘between-chunk transitions’. These two kinds of transitions emulate the varying transitional probabilities within and between words in language.

Finally, certain combinations of stimuli are not allowed within the grammar (and are ‘ungrammatical’) and as such never occur within the exposure stimuli (see below for details of the exposure and testing methodology) but occur in the ungrammatical sequences in the testing phase. Thus, these stimuli combinations have a TP equal to 0.0. These mimic the ungrammatical phrases in language.

Stimulus Sequences

The stimuli used in this experiment consisted of black shapes on a white background. We used eight such shapes as stimuli (A-H in Figure 1). They were abstract symbols which have been used previously in AGL studies (Fiser & Aslin, 2001). All the stimuli were of the same dimensions (200 pixels x 200 pixels), and they formed the vocabulary of the artificial grammar.

We needed grammatical and ungrammatical sequences in the various tasks of the experiment to test how and what people learned during the experiment. Eight unique grammatical sequences were generated from the artificial grammar (see Appendix for a list of the grammatical sequences). These sequences were six elements long with each element appearing only once in a given sequence. Each grammatical sequence had three within-chunk transitions. These within-chunk transitions were AB, CD, EF and GH and they were part of all eight grammatical sequences. Each grammatical sequence also had two between-chunk transitions with TP equal to either 0.67 (similar to *‘Happy birthday’*) or 0.33 (similar to *‘Happy baby’*). The high probability between-chunk transitions with TP equal to 0.67 were BC, DE, FG and HA and they were part of four high TP grammatical sequences. The low probability between-chunk transitions with TP equal to 0.33 were BG, DA, FC, HE and they were part of the rest of the four low TP grammatical sequences. The four high TP sequences were repeated twice as many times as the four low TP sequences in the task to generate the respective high and low transitional probabilities of the between-chunk transitions (see Appendix). This ensured that the participants were exposed to the high TP sequences a greater number of times than the low TP sequences, similar to the difference in exposure of pairs of syllables in language (*'Happy birthday'* vs *'Happy baby'*). ABCDEF and ABGHEF are examples of grammatical sequences generated from the grammar, the former being a high probability sequence while the latter being a low probability sequence as described above.

During the grammaticality judgement phase of the experiment (see the next section for more details), we wanted to test the kinds of regularities the participants had learnt. To test this, along with the grammatical sequences described above, we also needed ungrammatical sequences. We created three types of ungrammatical sequences, each of which violated our artificial grammar in three different ways (see Appendix for a list of the ungrammatical sequences). In the first type, we wanted to test if people had learned the within-chunk transitions. To test this, we created sequences that were mostly grammatical but with the exception of a violation in one of the within-chunk transitions in the sequence. These sequences had violations only in the within-chunk transitions, while the between-chunk transitions were grammatical and thus required knowledge of within-chunk transitions to judge them as ungrammatical. We created eight such ungrammatical sequences and called them 'within-chunk violation' sequences. For example, ABCFGH is one such within-chunk violation sequence where every transition between stimuli is legal except for the transition between C and F, which replaces the grammatical within-chunk CD. Similarly, we created the second type of ungrammatical sequence to test the participants' knowledge of the between-chunk transitions. In these sequences, only the between-chunk transitions were violated and thus required knowledge of between-chunk transitions to judge them as ungrammatical. We created eight such ungrammatical sequences and called them 'between-chunk violation' sequences. ABCDGH is an example of a between-chunk violation sequence, where every transition is allowed except for one violation in the transition between D and G, which is ungrammatical. Finally, we created the third type of ungrammatical sequence that had violations in the transitions of both 'between-chunks' and 'within-chunks'. We created eight such sequences which were called 'both violation' sequences since they violated both the transitions and could be judged as ungrammatical with the knowledge of either within-chunks or between-chunks. ABDCEF is an example of a both violation sequence which has two violations: the first violation, BD which is a between chunk violation and the second violation, DC which is a within chunk violation. Taken together, with eight unique ungrammatical sequences in three different types, we created 24 unique ungrammatical sequences to test the participant's knowledge of the different TPs in the grammaticality judgement phase (see Appendix for a list of all ungrammatical sequences in each type).

The Experimental Setup

Instructions about the experiment were given to the participant after which they were seated in front of a computer in the experiment room. The participants were explicitly told, "You will be shown a sequence of symbols on the screen. These sequences follow certain rules or patterns". They were also instructed to make any required responses by touching the touchscreen monitor.

The experiment was coded and run in MATLAB 2021a (Mathworks, USA) through the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007).

Procedure

Our experiment had three different tasks (Figure 1B) which are explained in detail in the following sections. The first task was the traditional Artificial Grammar Learning (AGL) task. It consisted of an ‘exposure phase’, where the participants were exposed to grammatical sequences, followed by a ‘grammaticality judgement’ testing phase in which participants judged sequences whether they were grammatical or not (see below for details). Participants completed three blocks of exposure and testing in the AGL task. This was followed by a ‘sequence completion’ task, where the participants had to complete partially completed grammatical sequences, and a ‘sequence generation task’, where the participants were instructed to produce complete sequences with no guidance (see below for details). Before the beginning of each task, the participants saw the task relevant instructions displayed on the screen.

Exposure phase: In the exposure phase of the AGL task, the participants were presented with grammatical sequences, generated from the grammar. Each element in the sequence was presented for a duration of 650ms with an inter-element interval of 130ms. After an inter-trial interval of 1s, the next exposure trial began. The elements were presented sequentially, one element after another. They were also presented in spatially different locations appearing left to right across the screen; thus, one element does not appear in the same location as the other elements in the sequence. The grammatical set of 12 sequences (see Stimulus Sequences) was repeated twice in a block, resulting in a total of 24 exposure trials. The order of the sequences was randomized across blocks and participants. Before the phase began, the participants were instructed only to observe the sequences and not to make any responses during this phase of the task.

Grammaticality Judgement (GJ) phase: The exposure phase was followed by a grammaticality judgement (GJ) phase. In this phase, we tested the participants' knowledge of the grammar by asking them to identify whether sequences were grammatical or not. Each trial began with the presentation of a sequence, similar to the exposure phase. This was followed by a screen with a green check mark and a red cross. If they thought that the sequence followed the same patterns or rules as the exposure sequences, they were instructed to touch the green check mark. On the other hand, if they thought that the sequence did not follow the same patterns or rules as the exposure sequences, they were instructed to touch the red cross. Once they made the decision, after an inter trial interval of 1s, the next trial began. In each block of the grammaticality judgement phase, the participants were presented with 24 grammatical sequences and 24 ungrammatical sequences resulting in 48 trials (see Stimulus Sequences).

Forced and Choice: We also wanted to test if the participants had acquired implicit or explicit knowledge of the grammar. To test this nature of the knowledge, we used the forced/choice paradigm from the metacognition literature (Hampton, 2001). In this paradigm, the participants are given two kinds of trials: forced trials and choice trials (Fig 1C). On choice trials, once the sequence was presented, participants were given a choice of whether to take the trial or to skip the trial. If they chose to skip the trial, the trial ended and the next trial began after the inter trial interval. If they chose to take the trial, they were shown the green check mark and red cross to judge the grammaticality of the sequence, as described above. However, on forced trials, the participants were not given any choice and the sequence presentation was directly followed by grammaticality judgement. Of the 48 trials in the grammaticality judgement phase, there were 32 choice trials and 16 forced trials. We ensured that the number of forced and choice trials was equal in grammatical and ungrammatical conditions. The order in which the sequences were presented was randomized across blocks and participants. Table 1 shows the distribution of the trials across different trial conditions.

Feedback: The trials were scored based on the participant's performance. Every correct trial added 100 points to the total score of the participant while an incorrect trial resulted in the deduction of 200 points from the total score. A skipped trial resulted in neither addition nor a deduction of points from the total score. The differential weights of correct and incorrect scores were given to motivate the participant to skip the trial if they were unsure of their judgement. The participants were instructed on the scoring system before the experiment began. However, we did not want the participants to learn the grammar during the testing phase. Immediate feedback after every trial could have helped the participants to learn the grammar by knowing the trials they got correct and those that they did not. Thus, to ensure that they could not learn during the grammaticality judgement phase, they were not shown their scores after every trial, but were shown their accumulated score every six trials. This ensured that they could not learn the grammar while also keeping them motivated to skip trials they thought were difficult by knowing about their performance in the task.

Sequence Completion Task: After completing three blocks of AGL task, participants proceeded to the sequence completion task. In this task, the participants were shown partially completed grammatical sequences with empty positions, along with the eight possible stimulus elements on the screen (see Fig. 1C). They were instructed to select the stimulus elements that they believed would fill these empty positions to complete the sequence correctly. As the participant selected the stimulus to fill an empty position, the selected stimulus appeared in that empty position of the partially completed sequence on the screen. Once the participant had filled all the empty positions, the next trial began after an inter trial interval of 1s.

Sequences: Similar to the grammaticality judgement phase, we wanted to test the participants' knowledge of the 'within' and 'between' chunk transitions. To test this, we generated two different types of partially completed grammatical sequences (see Appendix for a complete list of these sequences). They differed in the positions that were left empty for the participants to fill with appropriate stimuli. In the first type of trial, one stimulus element of the within-chunk transitions was left empty. Thus, to fill this type of sequence with the correct stimuli, the participant must have knowledge of the 'within-chunks'. A_C_E_ is an example of a 'partially completed within trial' where the second element of the within-chunks, are left empty for the participant to fill. We created eight such partially completed within trials to test the participants' knowledge of the within-chunk transitions. In the second type of trial, one stimulus of the between-chunk transitions and one stimulus of the within-chunk transitions was left empty. We called these the 'partially completed both trials' since they tested the participants' knowledge of both the transitions. ABCD__ is an example of a partially completed both trial where the empty positions require the between-chunk transition knowledge of DE along with the knowledge of the within-chunk transition EF to get the sequence correct. We created 16 partially completed both trials that tested participants on their within-chunk and between-chunk transitions knowledge. Taken together, we created 24 partially completed sequences, with eight partially completed within trials and 16 partially completed both trials. These 24 sequences were repeated thrice in the sequence completion task to give a total of 72 trials (Table 1). The order in which the sequences were presented was randomized across participants.

Forced and Choice: Similar to the grammaticality judgement phase, to test the nature of the knowledge that the participants had learnt, we used forced/choice conditions in this task. On choice trials, the participants had the option of either skipping or attempting to fill each position of the sequence in a given trial. If the participants choose to skip an unfilled position, a white box appears on that empty position, and the participant moves on to select their choice of stimulus element for the next unfilled position of the sequence. The participant did not have a limitation on the number of times they could choose to skip an empty position, and thus had the option to skip all the empty positions of the sequence. On forced trials, the participants were not given the option to skip filling any empty positions of the sequence. Of the total 72 trials, 24 trials were forced trials and 48 trials were choice trials and the trial type (partially completed within and between trials) was counterbalanced across the conditions.

Feedback: Each correctly filled position of the sequence added 100 points to the participant's total score while an incorrectly filled position resulted in a deduction of 200 points from their total score. Similar to the grammaticality judgement phase, we did not want the participants to learn from their errors by giving them immediate feedback after every trial or after every filled position in the sequence. Thus, similar to the grammaticality judgement phase, the participants were shown their accumulated score only every six trials.

Sequence Generation Task: The sequence completion task was followed by the sequence generation task. In this task, the participants were shown completely unfilled sequences with six empty positions, along with the eight possible stimulus elements on the screen (see Fig. 1C). They were instructed to generate a sequence, by filling the six empty positions with stimuli that they thought would appropriately follow the same rules or patterns that was followed by the sequences they had encountered previously in the experiment. The participants performed eight trials where they had to generate complete grammatical sequences. The participants had to take all the trials and there were no choice trials in this task. They were not given any feedback about their performance.

Verbal Reports: After the completion of the three tasks of the experiment, the participants were asked to fill a questionnaire regarding the task. The questions were designed to quantify the subjective experience of the participant while doing the task. They were asked to rate the difficulty of the three tasks on a scale of 1-5 with 1 being very easy and 5 being very difficult. They were also asked to rate their confidence in their responses in the three tasks, on a scale of 1-5, with 1 being complete guess and 5 being confident. They were also asked how motivated they were to skip a trial when given a choice in the grammaticality judgement task and the sequence completion task, on a scale of 1-5, with 1 being not motivated and 5 being very motivated. They were then given all the eight stimuli that made the artificial grammar, and using those stimuli they were asked to freely report any patterns they thought that they observed in the sequences presented to them.

Table 1

Distribution of trials across different trial conditions in Experiment 1.

<table>
<tr><th>Phase</th><th>Number of trials</th><th>Grammaticality</th><th>Level of choice</th></tr>
<tr><td>Exposure (repeated 3 times)</td><td>24 trials</td><td>All Grammatical</td><td>-</td></tr>
<tr><td rowspan="4">Grammaticality Judgement (repeated 3 times)</td><td rowspan="4">48 trials</td><td rowspan="2">24 Grammatical</td><td>8 Forced</td></tr>
<tr><td>16 Choice</td></tr>
<tr><td rowspan="2">24 Ungrammatical</td><td>8 Forced</td></tr>
<tr><td>16 Choice</td></tr>
<tr><td rowspan="2">Sequence Completion</td><td rowspan="2">72 trials</td><td rowspan="2">All Grammatical</td><td>24 Forced</td></tr>
<tr><td>48 Choice</td></tr>
<tr><td>Sequence Generation</td><td>8 trials</td><td>-</td><td>-</td></tr>
</table>

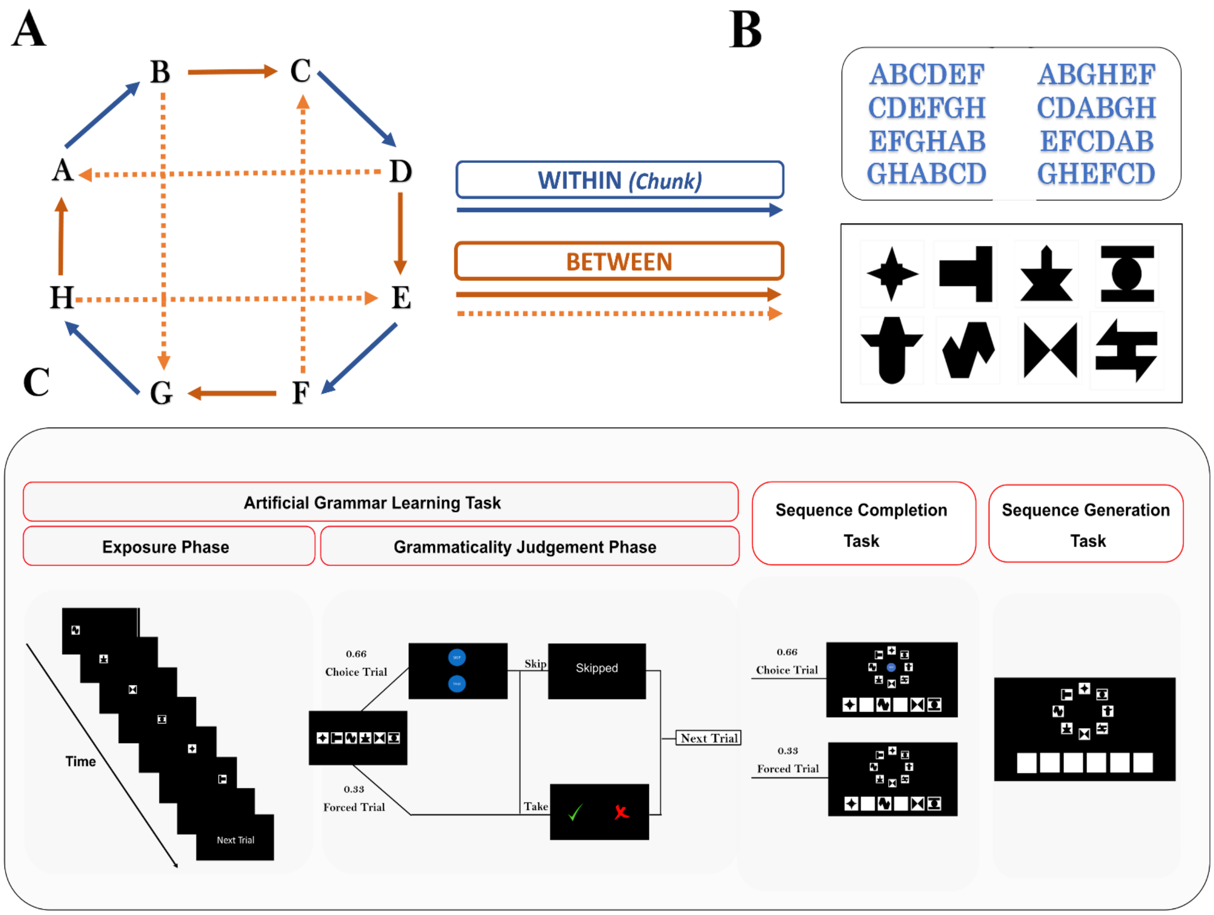

Figure 1:

(A) The artificial grammar used in Experiments 1 and 2. It consisted of eight elements A-H, whose order was dictated by different transition probabilities (TP). The blue arrows between the elements indicate a TP equal to 1 and these element pairs formed the within-chunks. The orange arrows between elements indicate a TP equal to 0.67 (solid orange) or 0.33 (dotted orange) and these element pairs formed the between-chunks. (B) Above - The eight unique grammatical sequences generated from the grammar and used in the exposure and the AGL phases of the task. Below – The eight visual stimuli as individual elements in the task. (C) The three tasks of the experiment. The first task is the Artificial Grammar Learning task, which consisted of an exposure phase and a grammaticality judgement phase. The grammatical sequences were presented, one element at a time in the exposure phase. This was followed by the grammaticality judgement phase where both, grammatical and ungrammatical sequences were presented sequentially after which the participant judged whether it was grammatical or not. 66% of the trials were choice trials and 33% of the trials were forced trials (see Procedure for more details). After three blocks each of exposure and GJ phase, sequence completion task followed. In this task, the participant was shown partially completed sequences with all possible stimuli on screen and were instructed to fill the empty positions with the correct stimuli. This was followed by sequence generation task, where the participant was instructed to generate complete sequences using the stimuli on screen and which they though followed the same pattern as the sequences they had observed earlier.

Results

Participants learnt the within-chunk transitions of the grammar in the AGL task:

To test if participants had learnt the artificial grammar, we looked at their performance in the grammaticality judgement (GJ) phase of the experiment. Across three blocks, participants performed at an average of 64% (Mean = 0.64, SEM = 0.02) and a one sample t-test comparing performance to chance (50%) showed that it was significantly above chance ($t_{(40)} = 7.95$, $p < 0.001$). A repeated measures ANOVA with Block as a within-measures factor (with three levels that indicate the three blocks) showed that they also performed better with time (Figure 2A: Main effect of Block, $F_{(1.5, 61.7)} = 31.22$, $p < 0.001$) indicating that they had learnt the grammar in the AGL task.

In the grammaticality judgement phase, the participants judged four different types of sequences - grammatical sequences and three types of ungrammatical sequences that had violations in the within-chunk transitions, between-chunk transitions and both transitions. We wanted to see if participants performed differently in these four different trial types using repeated measures ANOVA with Trial Condition as a factor (with four levels indicating the four types of trials). Participants showed different levels of performance in these trial conditions (Figure 2B: Main effect of Trial Condition, $F_{(3,120)} = 37.53$, $p < 0.001$). When averaged across three blocks, they performed significantly above chance in trials with sequences that were grammatical and in trials with sequences that had violations in within-chunk transitions and both transitions (Grammatical Sequences - Mean: 0.64, SEM: 0.02, one sample t-test: $t_{(40)} = 9.89$, $p < 0.001$; Ungrammatical Sequences with violations in within-chunk transitions – Mean: 0.67, SEM: 0.04, one sample t-test: $t_{(40)} = 4.66$, $p < 0.001$; Ungrammatical sequences with violations in both transitions – Mean: 0.63, SEM: 0.03, one sample t-test: $t_{(40)} = 4.35$, $p < 0.001$). The trials with these three types of sequences could be judged correctly only with the knowledge of the within-chunk transitions. Thus, these results show that the participants had learnt the within-chunk transitions. They also performed significantly below chance in trials with sequences that had violations in between-chunk transitions (Ungrammatical Sequences with violations in between-chunk transitions – Mean: 0.34, SEM: 0.03, one sample t-test: $t_{(40)} = -4.75$, $p < 0.001$). This type of trials could be judged correctly only if the participant had knowledge of the between-chunk transitions. The participants' performance in these trials suggests that they had not learnt the between-chunk transitions. This led them to systematically classify these trials as grammatical, thus leading to performance that was below chance (50%). Taken together, these results show that the participants had learnt the within-chunk transitions, but not the between-chunk transitions of the artificial grammar.

Participants showed evidence of explicit knowledge in verbal reports:

We had two metacognitive measures to assess the nature of the knowledge acquired in the AGL task, verbal reports and the difference in performance in forced-choice trials. For the verbal reports, participants were asked to report if they observed any patterns at the end of the experiment. From their reports, we coded whether the participants reported the within-chunk transitions (AB, CD, EF and GH) and between-chunk transitions (BC, DE, FG, HA, BG, DA, FC and HE). Of the 41 participants, 27 participants reported the within-chunk transitions and listed all four grammatical chunks (Figure 2C). However, no participant reported between-chunk transitions. This suggests that some participants acquired explicit knowledge of the within-chunk transitions while no participant had explicit knowledge of the between-chunk transitions.

The participants who reported the within-chunk transitions were labelled as 'reporters' and those who did not report the within-chunk transitions were labelled as 'non-reporters'. Reporters performed better than the non-reporters (Reporters - Mean: 0.68, SEM: 0.02; Non Reporters - Mean: 0.57, SEM: 0.02, $t_{(39)} = 3.002$, p = 0.002). We then analysed the performance between these two groups across the four different trial types using a repeated measures ANOVA with Trial Condition (four levels indicating the four trial types) and Reporting (with two levels indicating reporters and non-reporters) as factors. Results showed that there was a significant interaction between Trial Condition and Reporting suggesting that the reporters and non-reporters performance varied across the different trial types (Interaction between Trial Condition and Reporting: $F_{(3,117)} = 7.417$, $p < 0.001$). Post-hoc comparisons revealed that reporters performed better than non-reporters only in trials with grammatical sequences and in trials with sequences that had violations in the within-chunk transitions and both transitions (Figure 2D: Grammatical Sequences: Reporters – Mean: 0.776, SEM: 0.027, Non Reporters – Mean: 0.654, SEM: 0.038, p = 0.013; Ungrammatical Sequences with within-chunk transition violations: Reporters – Mean: 0.732, SEM: 0.040, Non Reporters – Mean: 0.538, SEM: 0.056, p = 0.008; Ungrammatical Sequences with between-chunk transition violations: Reporters – Mean: 0.299, SEM: 0.040, Non Reporters – Mean: 0.427, SEM: 0.055, p = 0.065; Ungrammatical Sequences with both transition violations: Reporters – Mean: 0.694, SEM: 0.034, Non Reporters – Mean: 0.514, SEM: 0.047, p = 0.004). These three trial conditions required the knowledge of the within-chunk transitions to judge the grammaticality of the sequence accurately. Taken together, these results suggest that the explicit knowledge of the within-chunks helped the reporters to perform better in those trials that required knowledge of the within-chunk transitions.

Participants did not show evidence of explicit knowledge in accuracy difference between forced and choice conditions:

The second metacognitive measure that we used to assess the knowledge acquired in the AGL task, was the difference in performance between forced and choice trials. If participants had acquired explicit knowledge of the grammar, then they should have selectively skipped those trials they thought were difficult and so would have performed better on choice trials compared to forced trials. We analysed the performance in the grammaticality judgement phase using repeated measures ANOVA with Choice (two levels with forced and choice trial types) as a factor. Results show that there was no significant main effect of Choice (Figure 4A; Performance averaged across three blocks: Forced trials – Mean: 0.64, SEM: 0.02; Choice trials – Mean: 0.643, SEM: 0.02; No significant main effect of Choice, $F_{(1,40)} = 0.021$, p = 0.885). We also tested whether participants showed this performance difference in any of the different trial types by running a repeated measures ANOVA with Trial Condition (with four levels indicating the four types of trials) and Choice (two levels) as two factors. There was no interaction between the factors Trial Type and Choice (Figure 3A: $F_{(3,87)} = 1.984$, p = 0.122) suggesting that there was no difference between forced and choice trials across all the trial types. Taken together, participants did not show signs of explicit knowledge as measured by the forced/choice paradigm.

However, as evidenced by the verbal reports, some of the participants had acquired explicit knowledge of the within-chunks. We wanted to test if these participants, the reporters, showed a difference in performance between forced and choice conditions. Using repeated measures ANOVA with Choice (two levels) as a factor, we analysed the performance of reporters and found no difference in performance between forced and choice conditions in reporters (Figure 3B: No significant effect of Choice, $F_{(1,26)} = 0.983$, p = 0.331). We also tested if there was an effect of the different trial conditions using repeated measures ANOVA with Trial Condition (four levels) as an additional factor along with Choice. There was a significant interaction between Trial Condition and Choice ($F_{(3,24)}$ = 3.381, p = 0.035), suggesting that the performance difference in forced and choice conditions varied according to the trial conditions. Further post hoc tests using Bonferroni corrections revealed that reporters performed better on choice trials than forced trials only in trials with grammatical sequences (Performance of Reporters on trials with Grammatical sequences: Forced – Mean: 0.747, SEM: 0.030; Choice – Mean: 0.799, SEM: 0.024, p = 0.019). They performed similarly in forced and choice trials in the rest of the three trial conditions (Performance of Reporters: Sequences that had violations in within-chunk transitions: Forced – Mean: 0.730, SEM: 0.043; Choice – Mean: 0.729, SEM: 0.045, p = 0.082; Sequences that had violations in between-chunk transitions: Forced – Mean: 0.297, SEM: 0.047; Choice – Mean: 0.270, SEM: 0.041, p = 0.112; Sequences that had violations in both transitions: Forced – Mean: 0.741, SEM: 0.044; Choice – Mean: 0.668, SEM: 0.040, p = 0.155). Taken together, even though verbal reports showed evidence of explicit knowledge of within-chunk transitions, when the reporters' performance was analysed based on forced/choice condition, the measure showed evidence of explicit knowledge only in the grammatical trial condition. Reporters did not show signs of explicit knowledge even in trials with sequences that had violations in the within-chunk transitions, which exclusively required the knowledge of only within-chunk transitions.

Participants did not show evidence of explicit knowledge in the Sequence Completion task:

After the completion of three blocks of the AGL task, participants did the sequence completion task where they filled partially completed grammatical sequences with stimuli, they thought fit the same pattern that they were exposed to earlier. To measure their performance in this task, we measured the percentage of the sequence that they filled that was grammatical. For example, if they had filled all the empty positions with the correct stimuli that together formed a grammatical sequence, that sequence was marked as 100% correct. However, if they had filled the incomplete sequence of 4 empty positions, with 2 correct stimuli and 2 incorrect ones, that sequence was marked as 50% correct. Participants, on average completed 73% of the sequence correctly (Fig 3C - Mean percentage of the sequence completed correctly - Mean: 0.73, SEM: 0.04) across all the trials. However, the task had three different trials that tested the participants on different TPs of the grammar, similar to the AGL task. Since the results from the AGL task showed that the participants had learnt only the within TPs, we predicted that in the Sequence Completion task, participants would perform better in trials that required only within knowledge than the other trials. A repeated measures ANOVA with Trial Condition (with two levels that included two types of trials, partially completed sequences within trials and partially completed both trials) and Block (with three levels indicating three blocks) as a factor showed a significant main effect of the Trial Condition (Fig 3C: Mean percentage of correct sequence: Partially completed within trial, Mean: 0.80, SEM: 0.05, Partially completed both trial, Mean: 0.69, SEM: 0.04; Effect of Trial Condition – $F_{(1,40)} = 25.485$, $p < 0.001$). Post hoc comparisons showed that participants performed better in partially completed within trials than in partially completed both trials, suggesting that participants used their knowledge of within-chunk transitions to perform well in those trials that required that knowledge.

We also tested whether the forced/choice metacognition measure showed evidence of explicit knowledge in the Sequence Completion task. A repeated measures ANOVA with Choice (with two levels that included forced and choice trial types) and Block (three levels) as a factor revealed no significant main effect of Choice (Fig 3C: Mean percentage of correct sequence:

Forced trials, Mean: 0.73, SEM: 0.040; Choice trials, Mean: 0.73, SEM: 0.044; Main effect of Choice – $F_{(1,40)} = 0.02$, p = 0.884) suggesting no evidence of explicit knowledge as measured by the forced/choice paradigm. Taken together, performance in the sequence completion task further adds to what was observed in the AGL task: participants had learnt the within-chunk transitions but showed no evidence of explicit knowledge as measured by the forced/choice paradigm.

Discussion

The results from the first experiment demonstrated that the participants had learnt the within-chunk transitions of the grammar (Fig 2B). Of all the participants, some participants showed evidence of having acquired explicit knowledge of the grammar by verbally reporting the within-chunk transitions (Fig 2C and D). However, these participants performed similarly in forced and choice trials including trials that required knowledge of only within-chunk transitions (Fig 2D), demonstrating no evidence of explicit knowledge as measured by the forced/choice paradigm.

Firstly, it is important to know whether the two metacognitive measures are assessing similar kinds of knowledge. These measures that we had used to assess knowledge might capture different aspects of knowledge, as proposed by Dienes and their colleagues (Dienes et al., 1991; Dienes & Scott, 2005). Knowledge that is acquired in an AGL task could be differentiated as structural and judgement knowledge. Structural knowledge is the knowledge of the structure of the artificial grammar (like rules or transition probabilities) that the participant acquires. This could be either implicit or explicit. Verbal reports test the explicitness of the structural knowledge by asking participants to report the structure of the artificial grammar. Accurate reporting in verbal reports suggest explicit structural knowledge while inaccurate or no reporting suggests implicit structural knowledge. This structural knowledge in turn, leads to another type of knowledge called judgement knowledge. Judgement knowledge is the knowledge state that is used by the participant to judge a sequence as grammatical or not in the grammaticality judgement phase. It is the feeling of *knowing* that a given sequence is grammatical or ungrammatical and it could be implicit or explicit. Structural knowledge, on the other hand, is the feeling of knowing *why* the sequence is grammatical or not. Explicit knowledge leads to explicit judgement knowledge, because if one knows *why* a sequence is grammatical by describing its structure, then they would also explicitly know that it is grammatical. However, implicit structural knowledge could lead to either implicit or explicit judgement knowledge – one could *know* a sequence is grammatical without knowing *why* they think so (explicit judgement knowledge) or one could *not know* if a sequence is grammatical while also not knowing *why* (implicit judgement knowledge).

In this experiment, we designed an AGL task with a new measure to assess knowledge. We propose that the performance difference between forced and choice trials is a measure of the judgement knowledge, and not structural knowledge. When a participant is given a choice to take or to skip a trial, they rely on the certainty of their judgement knowledge (‘*Do I know whether the sequence is grammatical or not*’) to make a decision which would be informed by their structural knowledge. If the participants had explicit judgement knowledge, then they would selectively skip trials that they are uncertain of resulting in a higher performance in choice trials than in forced trials. If the participants had implicit judgement knowledge, then they would not show a difference in performance between forced and choice trials.

The two metacognitive measures that we used in Experiment 1 measures two different knowledge states. Our results from verbal reports suggest that the reporters acquired explicit structural knowledge. This would in turn suggest that they should also possess explicit judgement knowledge. However, the forced/choice paradigm failed to provide evidence of explicit judgement knowledge.

One of the explanations for this could be that the participants might not have understood the paradigm. They might not have used the ‘Skip’ option appropriately and so did not skip those trials that they were uncertain about. This is evidenced by the performance of reporters in the trials that required only the knowledge of the within-chunk transitions. Even though they had accurate explicit structural knowledge of the within-chunk transitions, they were still not performing near ceiling (100%) in these trials that exclusively only required that knowledge (Fig 2D, Performance in trials with sequences that had violations in the within-chunk transitions). This suggests that they were still uncertain about their explicit structural knowledge which must have also resulted in uncertainty in their judgement knowledge resulting in some incorrect trials. However, in spite of this uncertainty, they did not skip trials that they were uncertain of which indicates that they might not have understood the paradigm. Thus, we designed our second experiment to address this issue by adding a positive control to the same task, to test if the participants used the skip option appropriately.

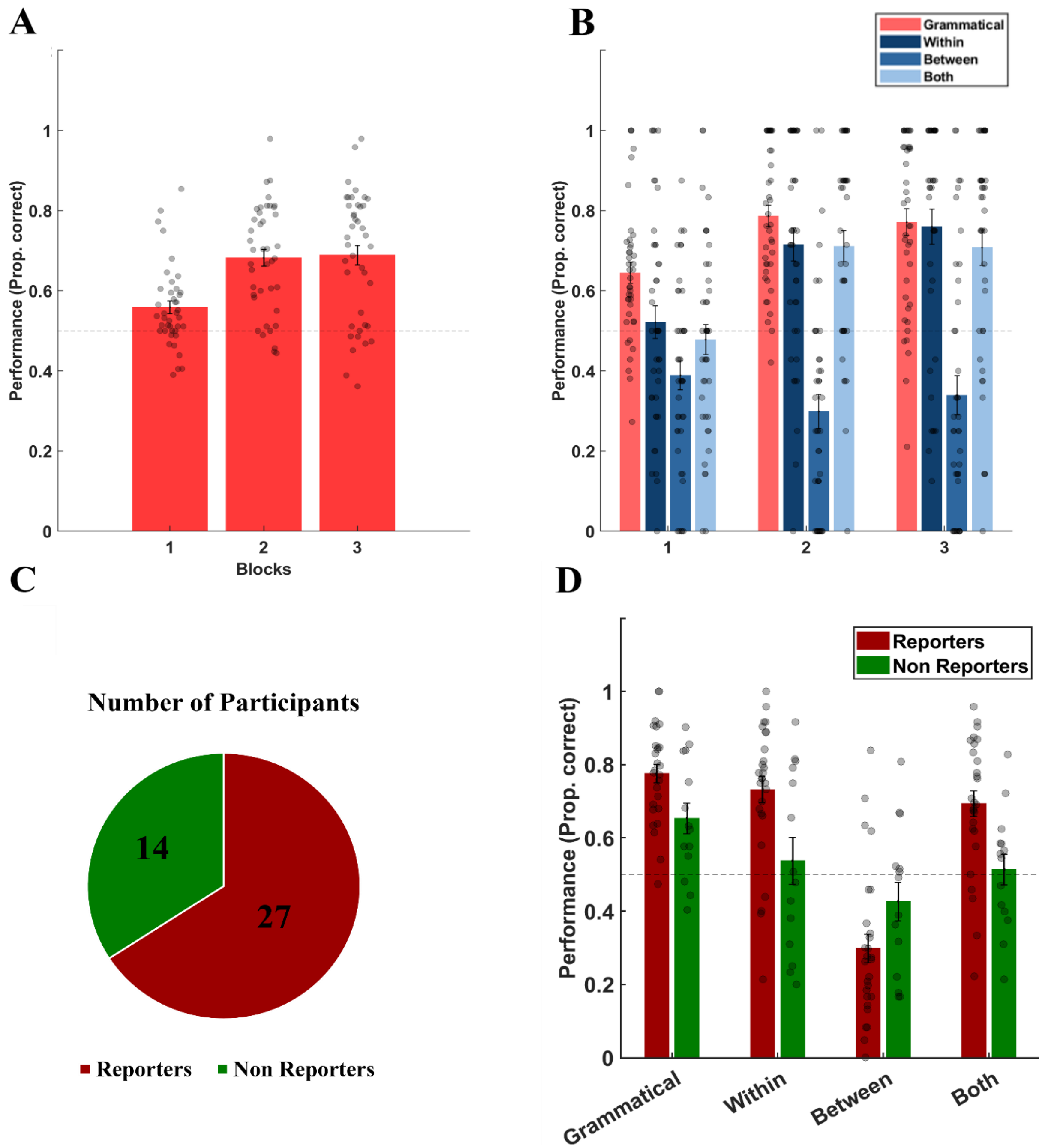

Figure 2:

(A) Mean performance in three blocks of the grammaticality judgement phase. (B) Mean performance in three blocks across trial conditions; From left: Red – Grammatical, Darkest blue – Within Ungrammatical, Blue – Between Ungrammatical, Lightest blue – Both Ungrammatical. (C): Pie chart showing the number participants who reported the within-chunks (maroon) and those who did not (green) (D). Performance of reporters (maroon) and non-reporters (green) across different trial conditions. The black circles correspond to data of individual participants for each graph. The dotted line represents performance at chance (50%).

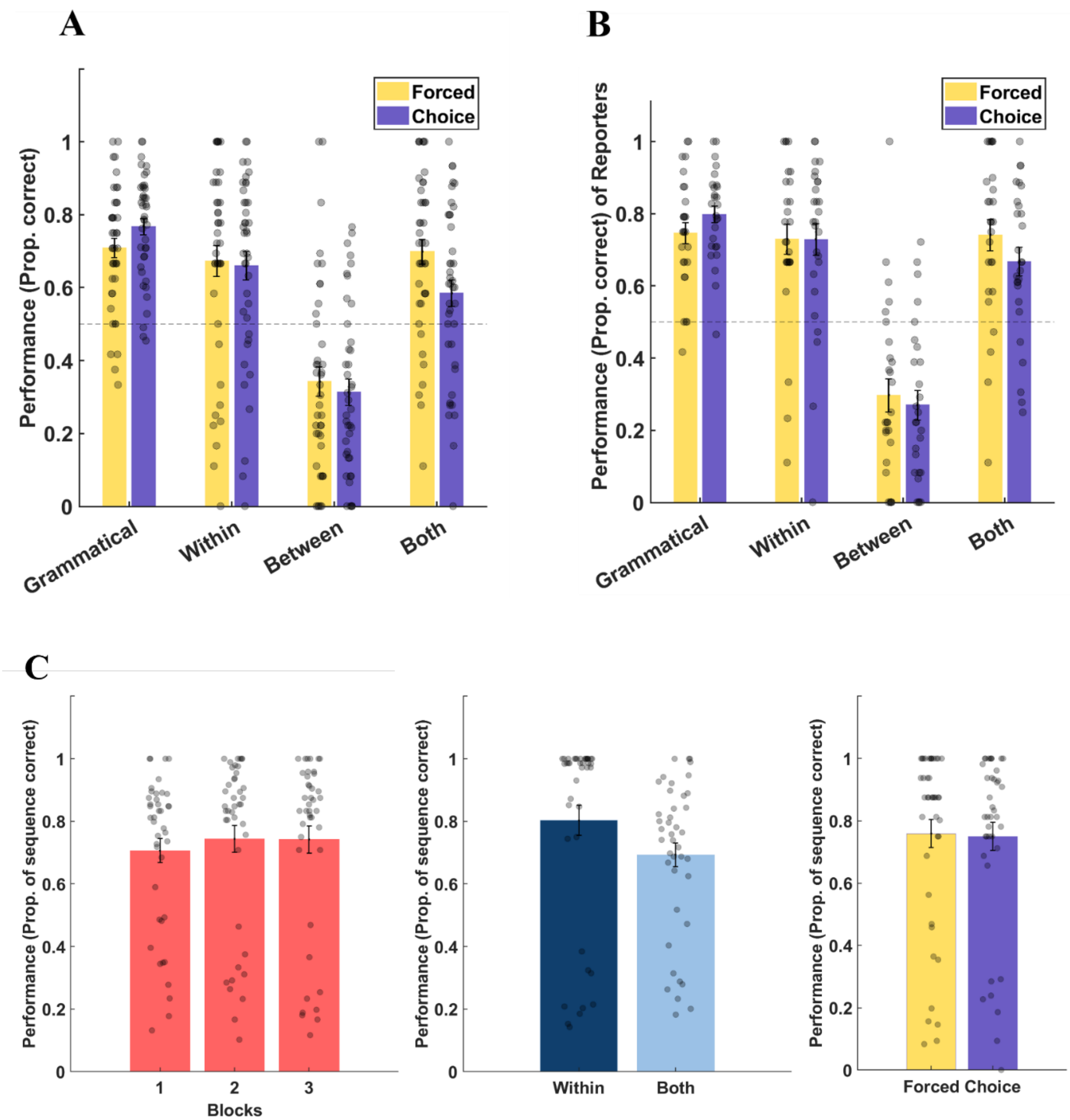

Figure 3:

(A) Mean performance of all participants in forced (purple) and choice (yellow) trials in different trial conditions in the grammaticality judgement phase. (B) Mean performance of the reporters in forced (purple) and choice (yellow) trials in different trial conditions in the grammaticality judgement phase. (C) Performance in Sequence Completion Task – First panel: Mean percentage of the sequence that was grammatically correct across participants in three blocks. Second panel: Performance in partially completed within trials (dark blue) and both trials (light blue) across three blocks. Third panel: Performance in forced trials (yellow) and choice trials (purple). The black circles correspond to data of individual participants for each graph. The dotted line represents performance at chance (50%).

Experiment 2

Results from Experiment 1 did not provide evidence of explicit judgement knowledge as measured by the forced/choice paradigm even though verbal reports showed evidence of explicit structural knowledge. Despite being uncertain about their judgement knowledge, the participants did not selectively skip trials that they were uncertain about. One of the reasons for this, could be that the participants might not have known how the task worked. Thus, to control for this possibility, we designed Experiment 2 with a positive control (explained in detail below) that tested whether the participants used the 'skip' option appropriately and therefore understood the experiment.

Participants

We recruited 43 participants (Mean age: 18.93 years; 11 males and 32 females) to participate in this experiment. All participants were Emory undergraduate students and were compensated for their participation through one research credit. The experiment was approved by Emory University IRB (ID: STUDY00003577).

Changes in Experiment 2

Positive control to test forced/choice trial conditions as a measure of explicit knowledge: As explained earlier, the results of Experiment 1 could be explained by the participants' not being able to understand how the task worked. Thus, they could have not skipped trials despite being uncertain of them. So, we wanted to increase this uncertainty by introducing some trials that were difficult in an evident way. To achieve this, we varied the presentation speed of the stimulus and introduced slow and fast trials, with the latter being more difficult than the former. Since increasing the difficulty would also increase the uncertainty of the participants' judgement, we hypothesized that if they were metacognitive about the difficulty of the trial, then they would skip more fast trials than the slow trials. If they skipped more fast trials appropriately, we would know that they understood the paradigm. Previous research has shown that stimulus presentation rate affects learning in visual AGL tasks (Conway & Christiansen, 2009; Turk-Browne et al., 2005). Even though participants learned the grammar in both slow and fast trials, they performed better in slow trials than in fast trials, suggesting that fast trials were more difficult than slow trials. Thus, we varied the difficulty of the trials with respect to the stimulus presentation speed to test whether the participants understood the paradigm.

We introduced slow and fast trials in both exposure and the grammaticality judgement phase. Each block of exposure and grammaticality judgement phase had 50% slow trials and 50% fast trials (Figure 4 and Table 2). On slow trials, the stimuli were presented for 650 ms with an inter-element interval of 130 ms same as the presentation speed of stimuli in Experiment 1. However, on fast trials, the elements were presented for 125 ms with an inter-element interval of 25 ms. The rest of the trial conditions of the grammaticality judgement phase was the same as in Experiment 1.

Addition of confidence ratings as another metacognitive measure: Since the metacognitive measures from Experiment 1 gave contrasting results about the nature of the acquired knowledge, we added confidence ratings in the grammaticality judgement phase as an additional measure of explicit judgement knowledge. Previous AGL studies have used confidence ratings as a measure of explicit judgement knowledge in AGL tasks (Dienes et al., 1991, 1995b). Typically, in these studies, the participants were instructed to rate their confidence after every trial in the grammaticality judgement task. Explicit judgement knowledge is inferred if the participants are more confident in trials in which they are correct and less confident in trials in which they are incorrect, resulting in a high performance-confidence correlation.

So, in the grammaticality judgement phase of Experiment 2, after the participant judged the grammaticality of the sequence, they were asked to rate the confidence of their response, on a scale of 1-100 (Figure 4). Once the participant rated their confidence, their rating was displayed on the screen and then the next trial began after an intertrial interval of 1s. Confidence ratings were asked from the participants on all forced trials and only on choice trials they chose to take.

Removal of the cumulative feedback: In Experiment 1, we gave cumulative feedback to increase the participants motivation to skip trials that they were unsure about. However, the participants did not appropriately skip trials even with cumulative feedback. Further, the addition of fast trials and confidence ratings increased the duration of the experiment. Thus, since feedback did not motivate them to skip the trials and to reduce the duration of the experiment, no feedback about their performance was given to the participants in Experiment 2.

Removing Sequence Completion and Sequence Generation: The sequence completion and sequence generation tasks did not provide any additional information regarding the nature of knowledge in Experiment 1. The results from Sequence Completion task were similar to the results from the AGL task. Further, the addition of another trial condition (slow/fast trials) and the inclusion of confidence ratings also increased the duration of the experiment. So, to minimize the experiment duration, we removed Sequence Completion and Sequence Generation tasks in Experiment 2.

Table 2

Distribution of trials across different trial conditions in Experiment 2.

Phase	**Number of trials**	**Grammaticality**	**Speed**	**Level of choice**
Exposure (repeated 3 times)	24 trials	All Grammatical	12 Fast	-
			12 Slow	-
Grammaticality Judgement (repeated 3 times)	48 trials	24 Grammatical	12 Fast	4 Forced
				8 Choice
			12 Slow	4 Forced
				8 Choice
		24 Ungrammatical	12 Fast	4 Forced
				8 Choice
			12 Slow	4 Forced
				8 Choice

A

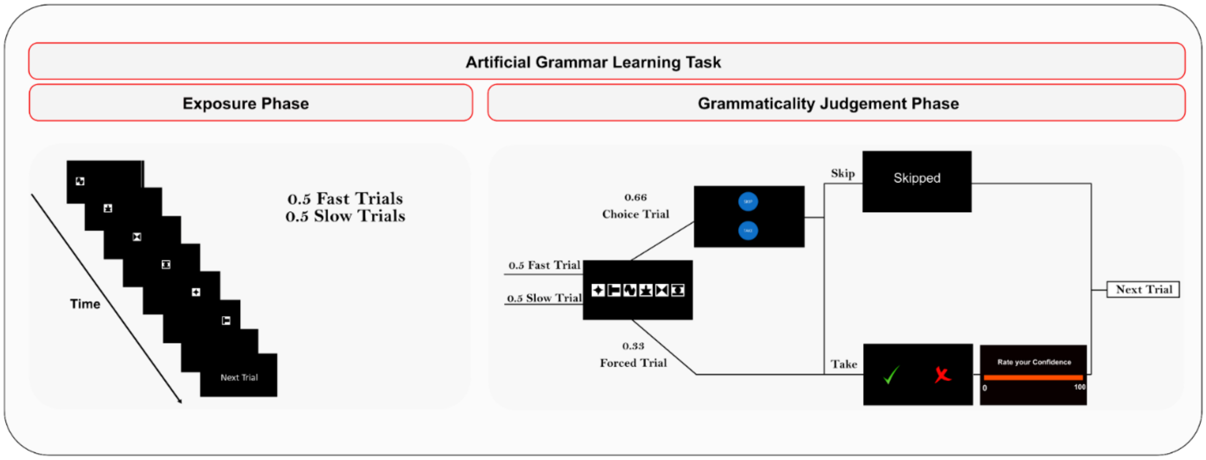

Figure 4:

The two tasks of Experiment 2. The Artificial Grammar Learning task consisted of an exposure phase and a grammaticality judgement phase. The exposure phase had 50% fast trials and 50% slow trials and, in these trials, the grammatical sequences were presented, one element at a time. Exposure was then followed by the grammaticality judgement phase where both, grammatical and ungrammatical sequences were presented sequentially after which the participant judged whether it was grammatical or not. Once the participant judges the sequences, they were asked to rate the confidence of their judgement on a scale of 1-100. This phase also had 50% fast trials and 50% slow trials. Further, 66% of the trials were choice trials and 33% of the trials were forced trials (see Table 2 for more details on trial distribution). The participants reported their confidence on all forced trials and on choice trials that they chose to take.

Results

Participants learnt the grammar:

In experiment 2, we observed a similar pattern of results to Experiment 1. Participants performed significantly above chance (Figure 5A; Performance – Mean: 0.57, SEM: 0.11, one sample ttest $t_{(42)} = 5.660$, $p < 0.001$) and they also performed better with time (Figure 5A: Repeated measures ANOVA: Main effect of Block - $F_{(2,84)} = 16.68$, $p < 0.001$) demonstrating that they had learnt the grammar. As in Experiment 1, a repeated measures ANOVA with Trial Condition as a factor (with four levels indicating the four types of trials in the grammaticality judgement phase) showed that the participants exhibited different levels of performance across different trial types (Figure 5B; Main effect of Trial Condition: $F_{(3,126)} = 46.19$, $p < 0.001$). Further, one-sample ttests revealed that their performance was better than chance only in trials with grammatical sequences and in trials with sequences that had violations in both transitions (Performance: Grammatical sequences – Mean: 0.712, SEM: 0.019, one sample t-test: $t_{(42)} = 10.983$, $p < 0.001$; Ungrammatical Sequences with within-chunk transition violations – Mean: 0.499, SEM: 0.027, one sample t-test: $t_{(42)} = -0.055$, $p = 0.478$; Ungrammatical Sequences with between-chunk transition violations – Mean: 0.317, SEM: 0.028, one sample t-test: $t_{(42)} = -6.574$, $p < 0.001$; Ungrammatical Sequences with both transition violations – Mean: 0.429, SEM: 0.024, one sample t-test: $t_{(42)} = -3.009$, $p = 0.002$). They also performed significantly below chance in trials with sequences that had violations in between-chunk transitions. Taken together, these results indicate that the participants had learnt the grammar. Even though they performed better in two types of trials, the results does not conclusively show whether the participants had acquired knowledge of within-chunk transitions or between-chunk transitions. Further, the performance on this experiment was lower than what was observed in Experiment 1 which could be due to the inclusion of fast trials that increased the difficulty of Experiment 2 and made it difficult for the participants to learn the grammar.

Participants showed evidence of explicit structural knowledge of within-chunk transitions:

Analysis of verbal reports allowed us to assess the nature of the structural knowledge acquired in the AGL task. Of the 43 total participants, 18 participants reported the within-chunk transitions while no participant reported the between-chunk transitions. The participants who reported the within-chunk transitions were labelled as 'reporters' and those who did not were labelled as 'non-reporters'. The proportion of reporters was also lower than what was observed in Experiment 1 which could again be because of the difficulty of Experiment 2.

To further analyse whether explicit structural knowledge helped the reporters to perform better in trials that exclusively required this knowledge, we did a repeated measures ANOVA with Trial Condition (with four levels indicating grammatical and three types of ungrammatical sequences) and Reporting (with two levels indicating reporters and non-reporters) as factors. Results showed that there was no main effect of Reporting suggesting that across all trial conditions, reporters performed comparably to non-reporters (Fig 5C: Performance of Reporters – Mean: 0.509, SEM: 0.02; Performance of Non-Reporters – Mean: 0.475, SEM: 0.017; Main effect of Reporting – $F_{(1,41)} = 1.832$, $p = 0.183$). However, there was a significant interaction between the two factors, Trial Condition and Reporting (Interaction between Trial Condition and Reporting: $F_{(1,123)} = 3.767$, $p = 0.013$) suggesting that the difference in the performance of reporters and non-reporters varied with respect to the trial condition. Post hoc comparisons using Bonferroni corrections revealed that the reporters performed better than the non-reporters in two types of trials: in trials with grammatical sequences and in trials with sequences that had violations in the within-chunk transitions (Figure 5D: Grammatical Sequences: Reporters – Mean: 0.758, SEM: 0.029, Non

Reporters – Mean: 0.679, SEM: 0.024, $p = 0.042$; Ungrammatical Sequences with within-chunk transition violations: Reporters – Mean: 0.574, SEM: 0.040, Non Reporters – Mean: 0.444, SEM: 0.034, $p = 0.017$; Ungrammatical Sequences with between-chunk transition violations: Reporters – Mean: 0.267, SEM: 0.042, Non Reporters – Mean: 0.353, SEM: 0.036, $p = 0.126$; Ungrammatical Sequences with both transition violations: Reporters – Mean: 0.439, SEM: 0.037, Non Reporters – Mean: 0.422, SEM: 0.031, $p = 0.729$). These two trial conditions required only the knowledge of the within-chunk transitions to judge the grammaticality of the sequence accurately. These results demonstrate that some participants acquired explicit structural knowledge of the within-chunk transitions of the grammar and this knowledge helped them to perform better than those who did not report only in trials that required this knowledge.

Participants showed evidence of explicit judgement knowledge in performance-confidence correlation:

In Experiment 2, we used an additional metacognitive measure, the correlation between participants' performance and confidence. If the participants had explicit knowledge of the grammar, then they would be more confident on trials that they were correct than on the trials that they were incorrect. Thus, their performance would be positively correlated with their confidence. To quantify this correlation, we measured a point biserial correlation between each participant's confidence in incorrect trials and their confidence in the correct trials. A positive correlation suggested that the participant was metacognitive about their knowledge of the grammar and thus had explicit knowledge of it. Across all the blocks, participants showed a positive correlation that was significantly higher than zero (One Sample t-test: Mean Correlation = 0.0937, SEM: 0.0178, $t_{(42)} = 5.255$; $p < 0.001$). A repeated measures ANOVA with Trial Condition as a factor (with four levels indicating the four types of trials in the grammaticality judgement phase) showed that the participants exhibited different levels of correlations across different trial types (Main effect of Trial Condition: $F_{(3,123)} = 11.412$, $p < 0.001$). Further, one-sample ttests revealed that their correlation was higher than zero only in trials with grammatical sequences and in trials with sequences that had violations in within transitions (Correlation between Performance and Confidence: Grammatical sequences – Mean: 0.245, SEM: 0.035, one sample t-test: $t_{(42)} = -7.253$, $p < 0.001$; Ungrammatical Sequences with within-chunk transition violations – Mean: 0.071, SEM: 0.05, one sample t-test: $t_{(42)} = -8.596$, $p < 0.001$; Ungrammatical Sequences with between-chunk transition violations – Mean: -0.169, SEM: 0.059, one sample t-test: $t_{(41)} = -11.344$, $p < 0.001$; Ungrammatical Sequences with both transition violations – Mean: -0.131, SEM: 0.521, one sample t-test: $t_{(42)} = -9.845$, $p < 0.001$). Taken together, these results showed that across different trial types the participants were more confident on trials they were correct and less confident on trials that they were incorrect, suggesting explicit judgement knowledge of the grammar.

We further tested whether reporters who had explicit knowledge of the within-chunk transitions showed a higher positive correlation than the non-reporters and whether the correlation differed with respect to the different trial conditions. We ran a repeated measures ANOVA with Reporting (two levels) and Trial Condition (four levels) as two factors. There was no significant effect of Reporting suggesting the reporters showed a similar correlation between their confidence and performance as the non-reporters (Fig 5C: Correlation between performance and confidence: Reporters – Mean: 0.083, SEM: 0.04; Non-Reporters – Mean: -0.005, SEM: 0.032; Main effect of Reporting – $F_{(1,40)} = 3.225$, $p = 0.080$). However, there was a significant interaction between the two factors, Trial Condition and Reporting (Interaction between Trial Condition and Reporting: $F_{(3,120)} = 8.291$, $p < 0.001$) suggesting that the difference in the correlation between reporters and non-reporters varied with respect to the different trial conditions. Post hoc comparisons using Bonferroni corrections revealed that the reporters showed higher correlation than the non-reporters in trials that had grammatical sequences and sequences that had violations in both transitions. (Figure 6: Correlation between performance and confidence: Grammatical Sequences: Reporters – Mean: 0.360, SEM: 0.050, Non Reporters – Mean: 0.157, SEM: 0.043, $p = 0.004$; Ungrammatical Sequences with within-chunk transition violations: Reporters – Mean: 0.171, SEM: 0.076, Non Reporters – Mean: -0.014, SEM: 0.065, $p = 0.072$; Ungrammatical Sequences with between-chunk transition violations: Reporters – Mean: -0.349, SEM: 0.083, Non Reporters – Mean: -0.034, SEM: 0.072, $p = 0.007$; Ungrammatical Sequences with both transition violations: Reporters – Mean: 0.152, SEM: 0.075, Non Reporters – Mean: -0.127, SEM: 0.065, $p = 0.008$). They also showed significantly lower correlation than non-reporters in trials that required knowledge of the between-chunk transitions suggesting that they were consistently incorrect on these trials while also being highly confident that they were correct. Taken together, these results suggest that across all participants, participants showed a positive correlation suggesting evidence of explicit judgment knowledge of the grammar as measured by the performance-confidence correlation measure. Further, explicit structural knowledge of the within-chunk transitions did not help the reporters to better access their explicit judgement knowledge than non-reporters.

Validation of the forced/choice paradigm:

Before assessing the nature of the knowledge that the participants had acquired through the forced/choice paradigm, we wanted to validate the paradigm through the positive control condition we had added in Experiment 2. We varied the speed of the stimulus presentation as a positive control to test if participants were metacognitive about the difficulty of the trial with respect to the stimulus presentation speed. If the participants thought that the fast trials were more difficult than the slow trials, then they would skip more fast trials, and their confidence would be lower in fast trials. Participants performed significantly better in slow trials than in fast trials (Performance in Fast trials - Mean: 0.528, SEM: 0.011; Slow trials – Mean: 0.594, SEM: 0.017, two sample ttest, $p < 0.001$). Participants also skipped fast trials more than the slow trials (Proportion of trials skipped: Fast: 0.355, SEM: 0.039; Slow: 0.208 SEM: 0.032, two sample ttest, $p < 0.001$). In the trials that they took (either forced or when given a choice) they were more confident in fast trials than in slow trials (Confidence – Fast: 0.471 SEM: 0.025, Slow – 0.581 SEM: 0.026, two sample ttest, $p < 0.001$). Taken together, these results suggest that the participants chose to skip or take the trial based on their perceived difficulty of the trial. They also rated those trials they perceived as difficult as less confident than those trials they did not perceive as difficult, suggesting that the participants were metacognitive about the difficulty of the trial and thus understood how the task worked.

Participants did not show evidence of explicit judgement knowledge through the performance difference between forced and choice trials:

Since the participants showed metacognition about the difficulty of the trial with respect to the stimulus presentation speed, we wanted to test if the measure could test the explicitness of the judgement knowledge of the participants. If participants had acquired explicit judgement knowledge of the grammar, then they would selectively skip trials that they are uncertain about resulting in higher performance on choice trials than on forced trials. A repeated measures ANOVA with two factors, Choice (two levels indicating forced and choice) and Trial Condition (four levels indicating trials with grammatical sequences and four types of ungrammatical sequences) showed that participants performed similarly on both forced and choice trials (Mean performance averaged across three blocks: Forced – 0.562 SEM: 0.013, Choice – 0.573 SEM: 0.016, p = 0.498; No significant main effect of Choice, $F_{(1,40)} = 0.021$, p = 0.885). There was also no interaction between these factors (Interaction between Choice and Trial Condition: $F_{(3,87)} = 1.984$, p = 0.122), suggesting that the participants performed similarly on forced and choice conditions across all trial conditions. Taken together, these results indicate that participants did not show signatures of explicit judgment knowledge as measured by the forced/choice paradigm across trial conditions.

We further wanted to test if reporters who has explicit structural knowledge of the within-chunk transitions performed better in choice trials than in forced trials in trial conditions that exclusively only require that knowledge. We tested the performance of reporters through a repeated measures ANOVA with two factors, Choice (two levels indicating forced and choice) and Trial Condition (four levels indicating trials with grammatical sequences and four types of ungrammatical sequences). We found no difference in performance between forced and choice conditions in reporters (Figure 6B: No significant main effect of Choice, $F_{(1,17)} = 0.892$, p = 0.358). However, there was a significant interaction between Trial Condition and Choice ($F_{(3,51)} = 3.716$, p = 0.017), suggesting that the performance difference in forced and choice conditions varied according to the trial conditions. Further post hoc tests using Bonferroni corrections revealed that reporters performed better on choice trials than forced trials only in trials with grammatical sequences (Fig 6B: Performance of Reporters on trials with Grammatical sequences: Forced – Mean: 0.713, SEM: 0.038; Choice – Mean: 0.811, SEM: 0.023, p = 0.013). They also performed worse on choice trials than forced trials in trials with sequences that had violations in the between-chunk transitions (Performance of Reporters on trials with sequences that had violations in between-chunk transitions: Forced – Mean: 0.337, SEM: 0.048; Choice –

Mean: 0.177, SEM: 0.032, p = 0.001) suggestive of the participants consistently choosing to take the between trial and incorrectly judging it as grammatical. They performed similarly in forced and choice trials in the other two trial conditions, sequences with violation in within-chunk transitions and in both transitions (Performance of Reporters: Sequences that had violations in within-chunk transitions: Forced – Mean: 0.552, SEM: 0.057; Choice – Mean: 0.586, SEM: 0.052, p = 0.639; Sequences that had violations in both transitions: Forced – Mean: 0.456, SEM: 0.047; Choice – Mean: 0.386, SEM: 0.044, p = 0.339). Thus, even though verbal reports showed evidence of explicit structural knowledge of within-chunk transitions, when the reporters' performance was analysed based on forced/choice condition, the results does not conclusively show that the participants had explicit judgement knowledge of the within-chunk transitions.

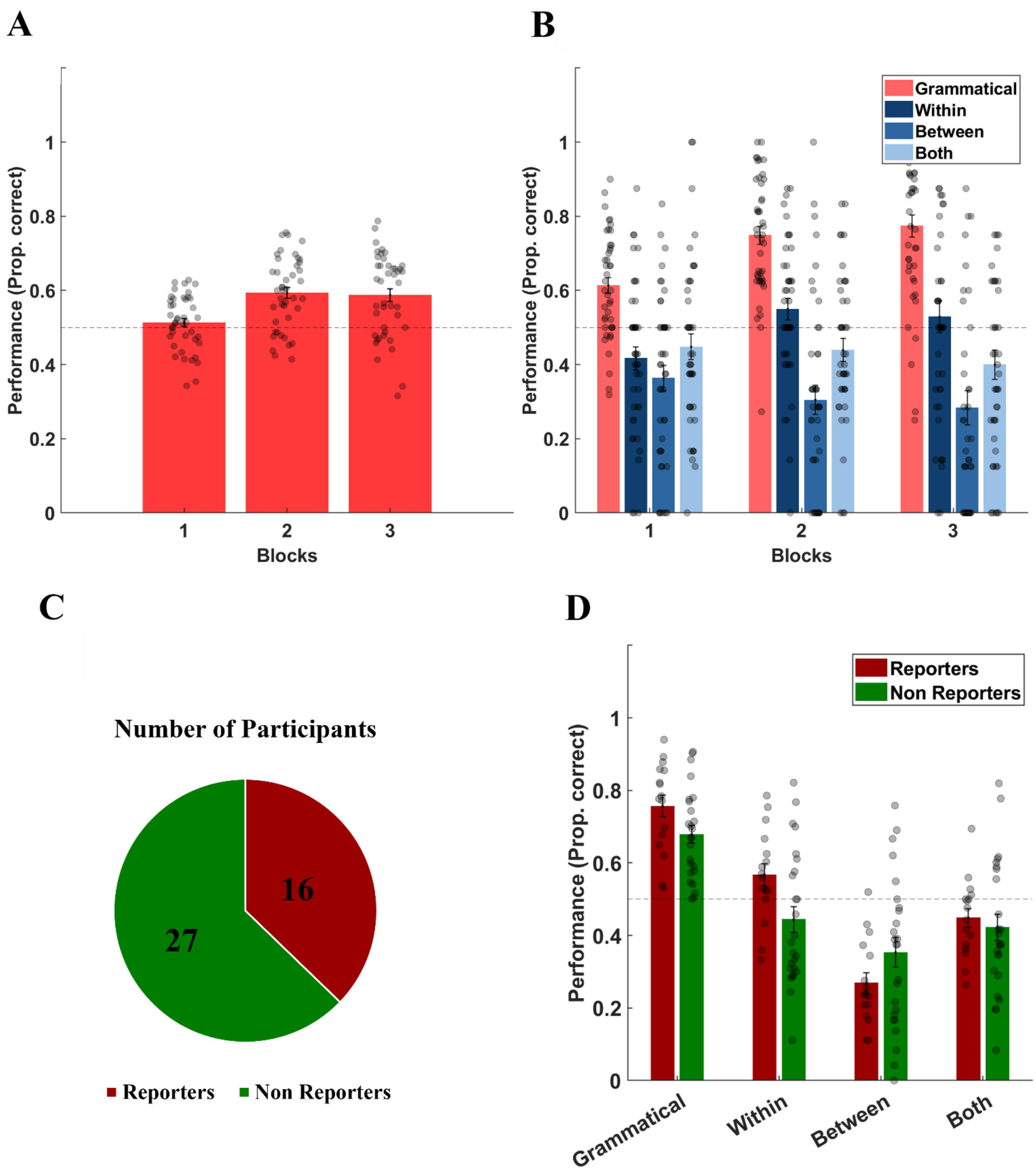

Figure 5:

(A) Mean performance in three blocks of the grammaticality judgement phase. (B) Mean performance in three blocks across trial conditions; From left: Red – Grammatical, Darkest blue – Within Ungrammatical, Blue – Between Ungrammatical, Lightest blue – Both Ungrammatical. (C): Pie chart showing the number of participants who reported the within-chunks (maroon) and those who did not (green) (D). Performance of reporters (maroon) and non-reporters (green) across different trial conditions The black circles correspond to data of individual participants for each graph. The dotted line represents performance at chance (50%).

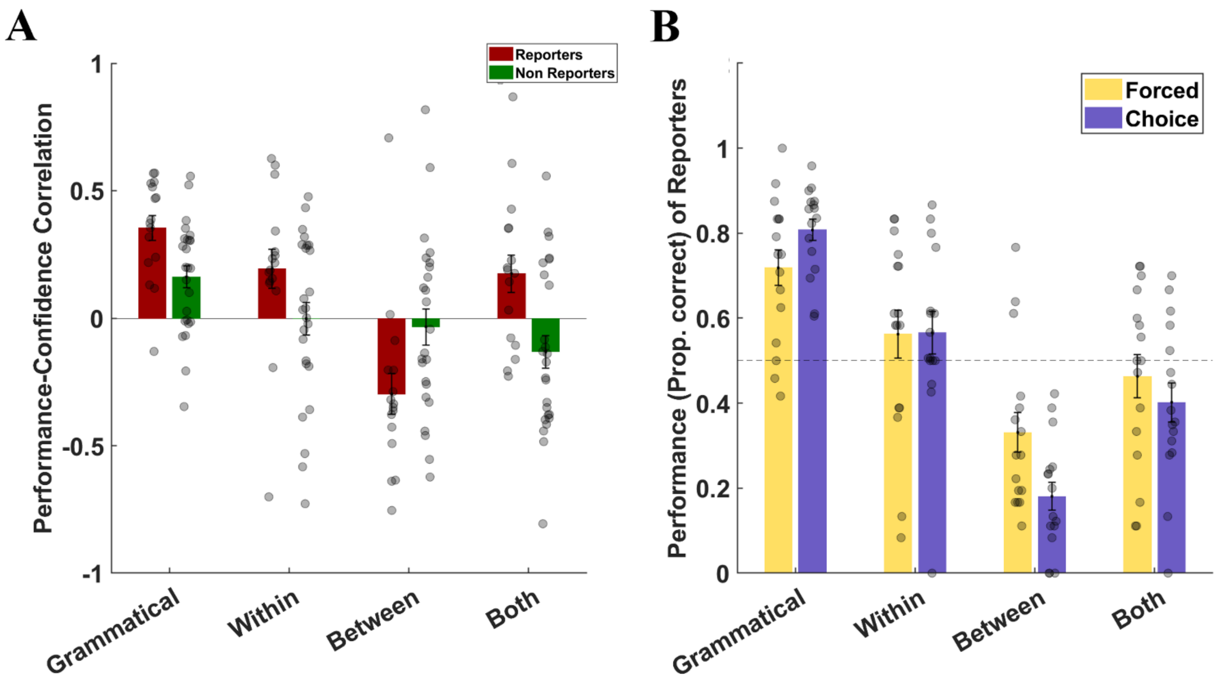

Figure 6:

(A) Correlation Coefficient between performance and confidence of reporters (maroon) and non-reporters (green) across the different trial conditions, averaged across three blocks. (B) Mean performance of the reporters on forced (yellow) and choice (purple) trials across different trial conditions, averaged across three blocks. The black circles correspond to data of individual participants for each graph. The dotted line represents performance at chance (50%).

Discussion

We used three metacognitive measures to assess the nature of knowledge that the participants had acquired from the AGL task in Experiment 2. The results indicate that the participants had learnt the grammar and that a subset of them had acquired explicit structural knowledge of the within-chunk transitions. Higher correlation between confidence and performance, especially in trials that required the knowledge of within-chunk transitions also provided evidence of explicit judgement knowledge in these participants. However, they did not show conclusive evidence of explicit judgement knowledge in forced and choice trials. Taken together, the results from these three metacognitive measures present contrasting evidence to the nature of the knowledge that the participants had acquired.

One of the reasons for these contrasting results of explicit judgement knowledge could be because of the way uncertainty influences these two measures. Similar to Experiment 1, results show that reporters with explicit structural knowledge of the within-chunk transitions did not perform at 100% even in those trials that exclusively only required knowledge of the within-chunk transitions (Fig 5D). This suggests that they had some uncertainty in their explicit structural knowledge and therefore also in their judgement knowledge. This uncertainty of the judgement knowledge was captured by the performance-confidence correlation measure by allowing the participants to rate their confidence as low. But the forced/choice could have failed to capture this uncertainty. This could be because of the difference in the range of the decision space of the forced/choice paradigm that had only two options of ‘take’ or ‘skip’ the trial and confidence ratings that had a scale of 1-100, thus having a larger decision space. In the forced/choice paradigm, due to its narrow range of possible decisions, the participants could have found it hard to map their subjective uncertainty on to their response (to take or to skip the trial). For example, a given participant, under same subjective levels of uncertainty, could have chosen to take some trials and skip other trials despite being equally uncertain at both times. This could be because they could not consistently map their subjective feeling of uncertainty with the limited range of the decision space. On the other hand, the confidence ratings offered the participant a much wider range of the decision space to map their uncertainty. Thus, the participants could have accurately mapped their subjective levels of uncertainty to their response (rating their confidence), leading in turn to an accurate measure of their judgement knowledge. Future experiments with either narrowing the range of confidence ratings to binary scale or a 5-point scale, equivalent to that of the forced/choice measure, could help address whether the granularity of the decision space could be a factor in accurate assessments of judgement knowledge.

Alternatively, the uncertainty of the participants could have been overridden by the explicit structural knowledge in the forced/choice paradigm. The decision to take or to skip the trial was made by the participants before they responded with their grammaticality judgement. Thus, the decision to skip or to take the trial, could have been influenced by a mix of their structural knowledge (‘*Do I know why I know the grammaticality of the sequence*’) and their judgement knowledge (‘*Do I know the grammaticality of this sequence*’), especially if they had explicit structural knowledge. If they had explicit structural knowledge (i.e., if the answer to the above posed questions are positive), then since they knew that they possess explicit knowledge, be it accurate or inaccurate, they could have felt motivated to use that explicit knowledge by choosing to take trials that they are uncertain of. This would result in inappropriate use of the skip option and thus an inaccurate assessment of the judgement knowledge due to the influence of structural knowledge. However, the performance-confidence correlation would have been uninfluenced by the structural knowledge because the participants rate their confidence after they had judged the sequence. Thus, they would have used only their judgement knowledge without the influence of structural knowledge while rating their confidence resulting in accurate assessment of the judgement knowledge.

Future experiments could test these measures in conditions that would produce implicit structural knowledge to test these explanations. One could change the complexity of the grammar since a more complex grammar could result in implicit structural knowledge (Reber, 1993). Since implicit structural knowledge could not influence the forced/choice accuracy difference, one would expect to see a difference in performance if the participants had acquired explicit judgement knowledge. Other task variables like the introduction of a cover task during the exposure phase could also be an interesting next step to assess these metacognitive measures under different experimental conditions.

References

Adams, A., & Santi, A. (2011). Pigeons exhibit higher accuracy for chosen memory tests than for forced memory tests in duration matching-to-sample. *Learning and Behavior*, *39*(1), 1–11. https://doi.org/10.1007/S13420-010-0001-7/FIGURES/7

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78. https://doi.org/10.1016/J.JML.2015.04.004

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Channon, S., Shanks, D., Johnstone, T., Vakili, K., Chin, J., & Sinclair, E. (2002). Is implicit learning spared in amnesia? Rule abstraction and item familiarity in artificial grammar learning. *Neuropsychologia*, *40*(12), 2185–2197. https://doi.org/10.1016/S0028-3932(02)00037-4

Christiansen, M. H. (2019). Implicit Statistical Learning: A Tale of Two Literatures. *Topics in Cognitive Science*, *11*(3), 468–481. https://doi.org/10.1111/tops.12332

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, *21*(4), 561–580. https://doi.org/10.1080/09541440802097951 de Vries, M. H., Barth, A. C. R., Maiworm, S., Knecht, S., Zwitserlood, P., & Flöel, A. (2010). Electrical stimulation of Broca's area enhances implicit learning of an artificial grammar. *Journal of Cognitive Neuroscience*, *22*(11), 2427–2436. https://doi.org/10.1162/jocn.2009.21385

Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995a). Unconscious Knowledge of Artificial Grammars Is Applied Strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1322–1338. https://doi.org/10.1037/0278-7393.21.5.1322

Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995b). Unconscious Knowledge of Artificial Grammars Is Applied Strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1322–1338. https://doi.org/10.1037/0278-7393.21.5.1322

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review 1997 4:1*, *4*(1), 3–23. https://doi.org/10.3758/BF03210769

Dienes, Z., Broadbent, D., & Berry, D. (1991). *Implicit and Explicit Knowledge Bases in Artificial Grammar Learning*. *17*(5), 875–887.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*(5), 735–755. https://doi.org/10.1017/S0140525X99002186

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research*, *69*(5–6), 338–351. https://doi.org/10.1007/S00426-004-0208-3/TABLES/6

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychologicl Science*, *12*(6), 499–504.

Fitch, W. T., & Hauser, M. D. (2004). Computational Constraints on Syntactic Processing in a Nonhuman Primate. *Science*, *303*(5656), 377–380. https://doi.org/10.1126/science.1089401

Foote, A. L., & Crystal, J. D. (2007). Metacognition in the Rat. *Current Biology*, *17*(6), 551–555. https://doi.org/10.1016/J.CUB.2007.01.061

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109–135. https://doi.org/10.1016/S0010-0277(99)00003-7

Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(9), 5359–5362. https://doi.org/10.1073/pnas.071600998

Inman, A., & Shettleworth, S. J. (1999). Detecting metamemory in nonverbal subjects: A test with pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*(3), 389–395. https://doi.org/10.1037/0097-7403.25.3.389

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact Artificial Grammar Learning in Amnesia: Dissociation of Classification Learning and Explicit Memory for Specific Instances. *Psychological Science, 3*(3), 172–179. https://doi.org/10.1111/j.1467-9280.1992.tb00021.x

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *20*(1), 79–91. https://doi.org/10.1037//0278-7393.20.1.79

Milne, A. E., Petkov, C. I., & Wilson, B. (2018). Auditory and Visual Sequence Learning in Humans and Monkeys using an Artificial Grammar Learning Paradigm. *Neuroscience*, *389*, 104–117. https://doi.org/10.1016/J.NEUROSCIENCE.2017.06.059

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*(2), 85–117. https://doi.org/10.1016/j.cogpsych.2003.12.002

Pelucchi, B., Hay, J. F., & Saffran, J. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development*, *80*(3), 674–685. https://doi.org/10.1111/j.1467-8624.2009.01290.x

Perruchet, P. (2019). What Mechanisms Underlie Implicit Statistical Learning? Transitional Probabilities Versus Chunks in Language Learning. *Topics in Cognitive Science*, *11*(3), 520–535. https://doi.org/10.1111/tops.12403

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. https://doi.org/10.1016/j.tics.2006.03.006

Ravignani, A., Sonnweber, R. S., Stobbe, N., & Fitch, W. T. (2013). Action at a distance: Dependency sensitivity in a New World primate. *Biology Letters*, *9*(6). https://doi.org/10.1098/rsbl.2013.0852

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(1), 88–94. https://doi.org/10.1037/0278-7393.2.1.88

Reber, A. S. (1989). Implicit Learning and Tacit Knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235.

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

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and Explicit Learning: Individual Differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 888–896. https://doi.org/10.1037/0278-7393.17.5.888

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, *33*, 829–856. https://doi.org/10.1017/S0142716411000580

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, *107*(2), 479–500. https://doi.org/10.1016/J.COGNITION.2007.10.010

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, *69*, 181. https://doi.org/10.1146/ANNUREV-PSYCH-122216-011805

Shields, W. E., Smith, J. D., & Washburn, D. A. (1997). Uncertain Responses by Humans and Rhesus Monkeys (Macaca mulatta) in a Psychophysical Same-Different Task. *Journal of Experimental Psychology: General*, *126*(2), 147–164. https://doi.org/10.1037/0096-3445.126.2.147

Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R., & Erb, L. (1995). The Uncertain Response in the Bottlenosed Dolphin (Tursiops truncatus). *Journal of Experimental Psychology: General*, *124*(4), 391–408. https://doi.org/10.1037/0096-3445.124.4.391

Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition*, *18*(3), 733–745. https://doi.org/10.1007/s10071-015-0840-x

Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition 2003 31:7*, *31*(7), 1060–1071. https://doi.org/10.3758/BF03196127

Turesson, H. K., & Ghazanfar, A. A. (2011). Statistical learning of social signals and its implications for the social brain hypothesis. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *12*(3), 397–417. https://doi.org/10.1075/IS.12.3.02TUR/CITE/REFWORKS

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945. https://doi.org/10.1162/JOCN.2009.21131

Wilson, B., Smith, K., & Petkov, C. I. (2015). Mixed-complexity artificial grammar learning in humans and macaque monkeys: Evaluating learning strategies. *European Journal of Neuroscience*, *41*(5), 568–578. https://doi.org/10.1111/ejn.12834

Yang, J., & Li, P. (2012). Brain Networks of Explicit and Implicit Learning. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0042993

Appendix

I. Grammatical Sequences

High Probability (repeated twice)	**Low probability (repeated once)**
ABCDEF	ABGHEF
CDEFGH	CDABGH
EFGHAB	EFCDAB
GHABCD	GHEFCD

II. Ungrammatical Sequences – The violations in the within-chunk transitions are bolded and underlined while the violations of the between-chunk transitions are only bolded. The sequences with violations in both transitions have violation of both within-chunk transitions and between-chunk transitions.

Within	**Between**	**Both**
ABC<u>**F**</u>GH	ABCD**G**H	AB**D**<u>**C**</u>EF
CDE<u>**H**</u>AB	EFGH**C**D	CD**F**<u>**E**</u>GH
EFG<u>**B**</u>CD	CDAB**E**F	EF**H**<u>**G**</u>AB
GHA<u>**D**</u>EF	GHEF**A**B	GH**B**<u>**A**</u>CD
ABG<u>**F**</u>CD	CD**G**HAB	AB**H**<u>**G**</u>EF
EFC<u>**B**</u>GH	GH**C**DEF	CD**B**<u>**A**</u>GH
CDA<u>**H**</u>EF	AB**E**FCD	EF**D**<u>**C**</u>AB
GHE<u>**D**</u>AB	EF**A**BGH	GH**F**<u>**E**</u>CD

III. Sequence Completion Sequences – the two types of partially completed sequences.

Within	**Both**	**Both**
A_C_E_	ABCD_ _	AB _ _ EF
C_E_G_	EFGH_ _	CD _ _ GH
G_E_C_	_ _ ABCD	EF _ _ AB
E_C_A_	_ _ EFGH	GH _ _ CD
_F_H_B	_ _ GHEF	_ _ CD _ _
_H_B_D	_ _ CDAB	_ _ EF _ _
_D_B_H	CDAB _ _	_ _ GH _ _
_B_H_F	GHEF _ _	_ _ AB _ _