



Lack of Cross-Modal Effects in Dual-Modality Implicit Statistical Learning

Xiujun Li^{1,2†}, Xudong Zhao^{2†}, Wendian Shi^{2*}, Yang Lu¹ and Christopher M. Conway^{3,4*}

¹ School of Psychology and Cognitive Science, East China Normal University, Shanghai, China, ² Department of Psychology, School of Education, Shanghai Normal University, Shanghai, China, ³ NeuroLearn Lab, Department of Psychology, Georgia State University, Atlanta, GA, United States, ⁴ Neuroscience Institute, Georgia State University, Atlanta, GA, United States

OPEN ACCESS

Edited by:

Petko Kusev,
University of Huddersfield,
United Kingdom

Reviewed by:

Paulo Carvalho,
Carnegie Mellon University,
United States
Valerio Santangelo,
University of Perugia, Italy

*Correspondence:

Wendian Shi
swd_nx@163.com
Christopher M. Conway
cconway@gsu.edu

[†]These authors have contributed
equally to this work as co-first
authors.

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 24 September 2017

Accepted: 29 January 2018

Published: 27 February 2018

Citation:

Li X, Zhao X, Shi W, Lu Y and
Conway CM (2018) Lack
of Cross-Modal Effects
in Dual-Modality Implicit Statistical
Learning. *Front. Psychol.* 9:146.
doi: 10.3389/fpsyg.2018.00146

A current controversy in the area of implicit statistical learning (ISL) is whether this process consists of a single, central mechanism or multiple modality-specific ones. To provide insight into this question, the current study involved three ISL experiments to explore whether multimodal input sources are processed separately in each modality or are integrated together across modalities. In Experiment 1, visual and auditory ISL were measured under unimodal conditions, with the results providing a baseline level of learning for subsequent experiments. Visual and auditory sequences were presented separately, and the underlying grammar used for both modalities was the same. In Experiment 2, visual and auditory sequences were presented simultaneously with each modality using the same artificial grammar to investigate whether redundant multisensory information would result in a facilitative effect (i.e., increased learning) compared to the baseline. In Experiment 3, visual and auditory sequences were again presented simultaneously but this time with each modality employing different artificial grammars to investigate whether an interference effect (i.e., decreased learning) would be observed compared to the baseline. Results showed that there was neither a facilitative learning effect in Experiment 2 nor an interference effect in Experiment 3. These findings suggest that participants were able to track simultaneously and independently two sets of sequential regularities under dual-modality conditions. These findings are consistent with the theories that posit the existence of multiple, modality-specific ISL mechanisms rather than a single central one.

Keywords: implicit statistical learning, cross-modal learning, modality-specific, multimodal input, dual-modality

INTRODUCTION

Human learners show sensitivity to environmental regularities across multiple perceptual modalities and domains even without being aware of what is learned (Aslin and Newport, 2009; Emberson and Rubinstein, 2016). This ability, referred to as implicit statistical learning (ISL), is a ubiquitous foundational cognitive ability thought to support diverse complex functions (Guo et al., 2011; Thiessen and Erickson, 2015).

A current debate in this area of research concerns the mental representations resulting from ISL. The nature of these mental representations is important for revealing the characteristics of the mechanisms underlying ISL (Cleeeremans and Jiménez, 2002; Fu and Fu, 2006; Li and Shi, 2016). In the classic study of implicit learning, Reber (1967) demonstrated ISL in participants

who were exposed to letter strings generated from an artificial grammar. In these experiments, letter strings obeyed the overall rule structure of the grammar, being constrained in terms of which letters could follow which other letters. Participants not only showed evidence of learning this structure implicitly, but also could apparently transfer their knowledge of the legal regularities from one letter vocabulary (e.g., M, R, T, V, X) to another (e.g., N, P, S, W, Z) as long as the underlying grammar used for both was the same. This effect has been replicated many times, with transfer being demonstrated not just across letter sets (Shanks et al., 1997), but also across perceptual modalities (Tunney and Altmann, 2001). The transfer effects in artificial grammar learning (AGL) are usually explained by proposing that the learning is based on abstract knowledge, that is, knowledge that is not directly tied to the surface features or sensory input (Reber, 1989; Altmann et al., 1995; Shanks et al., 1997; Peña et al., 2002). An additional characteristic of ISL is that it occurs with perceptually diverse input, including linguistic stimuli, tone stimuli, visual scenes, geometric shapes, color stimuli, and motor responses (Saffran et al., 1999; Fiser and Aslin, 2002; Kemény and Lukács, 2011; Durrant et al., 2013; Goujon and Fagot, 2013; Guo et al., 2013). Importantly, the same ISL phenomenon appears to be observed regardless of the nature of the input patterns. Given that ISL occurs with perceptually diverse input, it is possible that what underlies ISL is a single, central mechanism that treats all types of input stimuli (e.g., tones, shapes, and syllables) as equivalent beyond the statistical structure of the input itself.

However, there is evidence contrary to this view, suggesting that ISL is not neutral to the input modality but rather is rooted in modality-specific, sensorimotor systems. First, demonstrations of transfer of knowledge does not necessarily mean that the acquired knowledge is amodal. What is learned may be the surface characteristics of the stimuli but a separate, higher-level process may form mappings between the different types of input, allowing above-chance performance with the new input (Redington and Chater, 1996). Consistent with this view, a recent study showed that transferring knowledge to a new stimulus set in an AGL paradigm required working memory resources; when memory resources were depleted using a dual-task manipulation, no transfer effects were observed despite learning of the regularities occurring (Hendricks et al., 2013). Second, although ISL can occur with different types of stimuli, this does not necessarily indicate that ISL is subserved by a single, domain-general mechanism that applies across a wide range of tasks, inputs, and domains. Instead, it is just as possible that there may exist multiple parallel subsystems, each relying on similar computational algorithms, which can process and learn the underlying structure in various stimuli, (e.g., Chang and Knowlton, 2004; Conway and Christiansen, 2005; Conway and Pisoni, 2008; Goujon and Fagot, 2013; Frost et al., 2015). For example, Chang and Knowlton (2004) found that ISL was sensitive to stimuli features, and the changes in fonts could affect the ISL performance of the letter strings. This finding appears to indicate that at least some of the learned knowledge is modality or stimulus specific. In addition, using vibration pulses, pictures, and pure tones as experimental materials, Conway and Christiansen (2005) compared tactile, visual, and auditory ISL

and found modality constraints affecting ISL across the senses, with auditory ISL showing better performance than both tactile and visual learning (see also Conway and Christiansen, 2006, 2009). Similarly, Emberson et al. (2011) presented visual and auditory input streams under different timing conditions (fast or slow presentation rates). The results showed that auditory ISL was superior to visual learning at fast rates, but the opposite was true at slower presentation rates, suggesting the existence of modality constraints affecting learning.

The studies reviewed to this point relied on comparisons across individual modalities. However, the perceptual environment is rarely limited to one modality or a single information stream (Stein and Stanford, 2008), and learners often face multiple potential regularities across modalities at the same time. Research in the area of multisensory integration suggests that to some extent, information can be processed separately and in parallel across different perceptual modalities. For instance, participants can monitor simultaneously visual and auditory inputs in different spatial locations without a behavioral deficit under conditions of divided attention (Santangelo et al., 2010). In addition, findings from working memory research suggests that when information is presented in both visual and auditory-verbal formats, the information is encoded separately and yet a facilitative effect is also observed in bimodal formats (e.g., audiovisual stimuli), leading to improved memory (Mastroberardino et al., 2008). Although these studies demonstrate the manner in which multimodal input streams are processed by attentional, perceptual, and memory mechanisms, it is currently unclear to what extent ISL can support such processing demands.

It is important therefore to explore the degree to which multimodal information streams are processed independently or are integrated together to support implicit learning (e.g., Sell and Kaschak, 2009; Cunillera et al., 2010; Mitchel and Weiss, 2010, 2011; Thiessen, 2010; Shi et al., 2013; Mitchel et al., 2014; Walk and Conway, 2016). The extent to which simultaneous multisensory input are processed independently rather than being integrated together provides perhaps the strongest support for the existence of multiple, modality-specific mechanisms of ISL. That is, learning of multiple input streams in parallel does not seem feasible for a single central learning mechanism; only if multiple learning mechanisms exist could parallel input streams be learned and represented independently of one another.

Conway and Christiansen (2006) assessed multistream learning in a series of three experiments. In Experiment 1, participants were exposed alternately with auditory sequences produced from one artificial grammar and visual sequences generated by a second grammar. In the test phase, new sequences were generated from each grammar; crucially, for each participant, all sequences from both grammars were instantiated only visually or auditorily. The results revealed that participants only endorsed a sequence as “grammatical” if the sensory modality matched the grammar that it was paired with during the learning phase. These findings suggest that ISL is closely bound to the input modality in which the regularities are presented, rather than operating at an abstract level. Johansson (2009) extended these findings by increasing the amount of exposure during the

learning phase, believing that this would be more likely to result in formation of abstract representations. The results were still consistent with stimulus-specific, not abstract, representations.

However, due to the crossover design used (Conway and Christiansen, 2006; Johansson, 2009), these two studies were not able to examine the learning of cross-modal sequences at the same time in a strict sense. That is, the visual sequences and auditory sequences were interleaved and alternated with one another rather than being presented concurrently. Therefore, in order to provide evidence that multiple sensory modalities can be used to learn sequential regularities simultaneously and independently, a different type of design is necessary.

In an initial study using a dual-modality design, Shi et al. (2013) presented participants with visual and auditory sequences simultaneously, examining the degree to which multimodal input sources are processed independently. They found that the participants could acquire the regularities presented simultaneously regardless of the grammatical rules being the same or different, and there were no significant differences between unisensory and multisensory conditions. That is to say, audiovisual presentation of the same regularities did not show an enhanced ISL effect one might expect if there was a single learning system integrating the perceptual inputs. Similarly, audiovisual presentation of different grammars at the same time also did not show an interference effect, suggesting that the learning of two sets of regularities presented in two different perceptual modalities can occur independently and in parallel.

In the Shi et al. (2013) experiments, the audiovisual sequences consisted of nonsense syllables and color pictures paired together. The absence of enhanced or decreased ISL during the cross-modal presentation conditions may be due to the nature of the stimuli. Specifically, Conway and Christiansen (2006) found that when two input streams following different grammatical rules were presented in separate perceptual dimensions (e.g., unfamiliar shapes and nonsense syllables), participants were able to demonstrate learning of the two grammars. However, when the two input streams containing different grammatical rules were presented in the same perceptual dimension (e.g., two different sets of nonsense syllables or two different sets of unfamiliar shapes), participants were not able to learn both sets of regularities (Conway and Christiansen, 2006). Therefore, the perceptual characteristics of the stimuli appear to play an important role in cross-modal ISL.

Based on these findings, we conjectured that cross-modal effects would be more likely to occur if the two input streams contained high perceptual overlap, namely, visual and auditory stimuli that referred to the same objects. Specifically, in this study, we used animal pictures as the visual stimuli and their names as the auditory stimuli. Using pictures of animals and their names presented concurrently should provide the strongest test of parallel learning mechanisms. That is, if learning of the grammatical patterns operates independently over visual and auditory input, then concurrent presentation of animal pictures with their auditory names should show no behavioral facilitation. On the other hand, if learners are representing the visual pictures and auditory names as a single perceptual event (e.g., a multimodal percept that includes the picture of the animal tied to

its name), then we might expect to observe better learning under concurrent compared to unimodal presentation. Similarly, if the animal pictures and animal names are presented in sequences generated from two different artificial grammars at the same time, this would be expected to result in an interference effect if learners are trying to integrate the pictures with the names. But if no interference is observed under such seemingly difficult learning conditions, this would be strong evidence for parallel, independent learning mechanisms.

The objective of this study therefore is twofold. First, we tested whether multistream, cross-modal ISL results in increased learning when the auditory and visual stimuli provide redundant information. Second, we explored whether multistream cross-modal ISL results in decreased learning when the auditory and visual stimuli provide conflicting information. The answers to these questions will ultimately allow us to determine whether the cross-modal patterns were processed independently within each modality or whether the perceptual modalities are integrated together during the learning process.

EXPERIMENT 1A: UNIMODAL VISUAL ISL

Visual ISL of a single modality was tested, and used to establish a baseline level of performance for comparison with dual-modality ISL in subsequent experiments.

Method

Subjects

Twenty-two Chinese graduate students with normal hearing and normal or corrected-to-normal vision were recruited from Shanghai Normal University via an advertisement (Age range = 23–29; Mean age = 25.75; Females = 13). The decision of the sample size of 22 was predetermined by priori power analysis, based on the G*Power package version 3.1.9.2, which indicated the sample size > 19 adequately makes a moderate experimental effect (Cohen's $d = 0.8$) being detectable with power = 0.9 at alpha = 0.05. None of them had ever participated in any type of cognitive experiments. This study was carried out in accordance with the recommendations of the Ethics committee of the Shanghai Psychological Society with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki, and they were paid for their participation. The protocol was approved by the Ethics committee of the Shanghai Psychological Society.

Materials

An artificial grammar (see **Figure 1**) (Christiansen et al., 2012) was used to produce a set of sequences containing between five to seven elements. Each letter of the grammar was mapped onto an animal vocabulary including tiger, lion, elephant, horse and goat. The grammar determined the order of sequence elements drawn from five different categories of stimulus tokens. Two categories, A and B, each contained a single token, a tiger (A) and a lion (B), respectively. The C category consisted of two tokens, a black elephant (C_1) and a gray elephant (C_2).

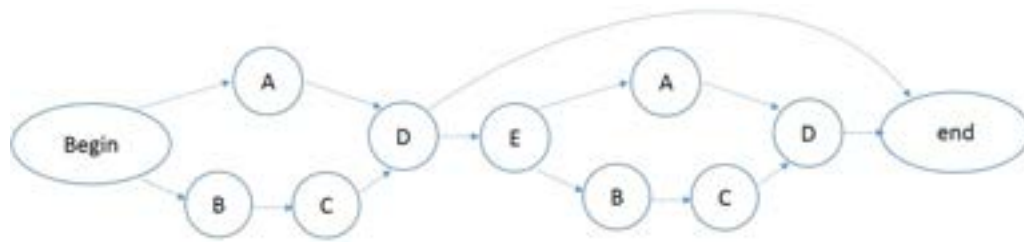


FIGURE 1 | Artificial grammar 1.

The D and E categories each contained three tokens, a white horse (D₁), a black horse (D₂), and a gray horse (D₃); and a white goat (E₁), a black goat (E₂), and a gray goat (E₃), respectively. There were a total of 10 tokens distributed over the ten stimuli.

A large number of sequences can be generated according to this artificial grammar. We used 48 legal sequences that were generated from the grammar for the acquisition phase (see Appendix A, Grammar 1). The test set (see Appendix B) consists of 80 novel legal and 80 illegal sequences. The illegal sequences each began with a legal element, followed by several illegal transitions and ending with a legal element. For example, the illegal sequence B–C₁–E₂–A–D₁–D₁ begins and ends with legal elements (B and D, respectively) but contains two illegal interior transitions.

A possible token sequence resulting from this artificial grammar could be A–D₁–E₂–B–C₁–D₂, which corresponds to a tiger, white horse, black goat, lion, black elephant, and black horse (see Figure 2).

Procedure

The experiment consisted of two phases: a learning phase and a testing phase. At the beginning of the acquisition phase, participants were told a story adapted from Rosas et al. (2010) but presented in Chinese: “Thomas travels along his country in a train carrying his circus. They arrived in a total of 144 cities. Whenever they arrive in a city, Thomas makes his animals perform. Next, you’ll see sequences of animals on the computer screen. They are the appearance order of animals performing at each city. Please remember these sequences and afterward you will take a test.”

The participants were not told that the sequences had been produced according to an artificial grammar. The sequences were presented one at a time to each participant. Each animal picture in the sequences lasted for 1000 ms with 300 ms inter-animal intervals. A 2000 ms pause occurred between the first and second sequences. Each of the 48 learning phase sequences was presented three times in random order for a total of 144 trials.

After the familiarization phase, participants were presented with the testing phase. The testing phase started with the following story: “The animals appearing in the orders you have observed were generated according to a set of complex rules that determined the order of the animals within each city. Thomas will travel along another country, so animals appearing in new orders have been produced. Some of these new orders conform to the same

rules as before, the others are different. Only the orders conforming to the same rules would be allowed to perform on stage. Next you will see the animals appearing in new orders. Your task is to determine if the orders conform to the same rules as before by pressing one of two buttons marked YES and NO without feedback. You need to respond quickly and accurately.” Then, the 160 test sequences were presented in random order to the participants. The presentation time of the test sequences was the same as that used during the acquisition phase.

After completing the experiment, the participants were debriefed about their explicit knowledge of the rules as well as any particular strategies that they might have used during both the training and testing phases.

Results and Discussion

The mean test accuracy in Experiment 1A was 122 out of 160 (76%), with a standard deviation of 24. A one-sample *t*-test indicated that performance for the unimodal visual ISL task was significantly above chance, $t(21) = 8.26$, $p < 0.001$, $d = 1.73$.

EXPERIMENT 1B: UNIMODAL AUDITORY ISL

Auditory ISL of a single modality was tested and used to establish a baseline level of performance for comparisons with dual-modality ISL in subsequent experiments.

Method

Subjects

Twenty-four new Chinese graduate students were recruited in the same manner as in Experiment 1A (Age range = 23–29; Mean age = 25.88; Females = 14). A parallel power analysis with Experiment 1A was applied, and yielded a result that a sum of 46 participants was acquired to detect a large condition effect across groups (Cohen’s $d = 1$) at level of power = 0.9 and alpha = 0.05. Thus, subject number was 24, generated by subtracting subject number used in Experiment 1A ($N = 22$) from 46.

Materials

The experimental materials were the same as Experiment 1A except that animal names were presented in the auditory modality instead of pictures. Animal names recorded by a female, native Chinese speaker. They were presented via headphones.

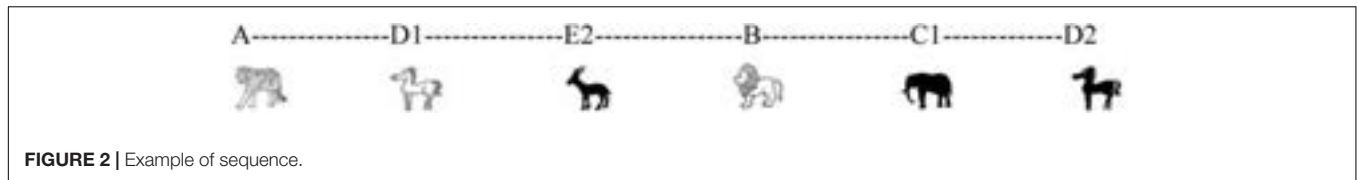


FIGURE 2 | Example of sequence.

Procedure

The procedure was identical to Experiment 1A, except that all references to animal pictures were replaced with sounds.

Results and Discussion

The mean test accuracy in Experiment 1B was 118 out of 160 (74%), with a standard deviation of 25. A one-sample *t*-test indicated that performance for auditory ISL was significantly above chance, $t(23) = 7.06$, $p < 0.001$, $d = 1.50$. There were no statistical differences between visual and auditory levels of learning, $t(44) = 0.59$, $p = 0.56$.

In both Experiments 1A and 1B, we asked participants whether they had based their judgments on specific rules or strategies. Most participants reported basing their responses merely on whether a sequence felt familiar or similar. These verbal reports suggest that they had very little explicit knowledge concerning sequence legality. Several of the participants reported that they made their judgments on the basis of a simple rule (e.g., “Certain animal combinations were not allowed,” or “If certain combinations of animals were the same as the learning phase, I said ‘yes.’” However, none of the participants was able to report anything specific that could actually help him or her make a test decision. On the basis of these verbal reports, there appears to be little evidence that the participants were explicitly aware of the distinction between legal and illegal sequences, despite their above chance performance.

The results of Experiments 1A and 1B demonstrate that participants acquired knowledge of both the visual and auditory regularities under standard unimodal conditions. Next, we presented sequences of stimuli in both visual and auditory modalities simultaneously under dual-modality conditions either using a single grammar (Experiment 2) or a different grammar for each modality (Experiment 3). The data obtained from Experiment 1 were used as baseline levels of performance to compare to levels of performance in the dual-modality conditions.

EXPERIMENT 2: DUAL-MODALITY, CROSS-MODAL ISL WITH ONE GRAMMAR

The primary goal of Experiment 2 was to examine the effect of redundant cross-modal information on ISL. More specifically, this experiment was designed to examine whether ISL would be enhanced when participants were presented with corresponding auditory and visual sequences that contained the same grammatical regularities. If there is no difference in learning between this type of cross-modal format and the

unimodal baseline conditions, then this would suggest that the sequential regularities in each perceptual domain was processed independently of the other. Instead, if behavioral facilitation is observed, this would suggest that the visual and auditory inputs were integrated together to form a unified cross-modal representation of the sequential regularities.

Method

Subjects

Fifty-two Chinese graduate students were recruited (Age range = 23–30; Mean age = 25.96; Females = 22). We excluded 2 subjects from analysis who failed to follow instructions. The priori power analysis showed, despite 2 participants’ data was excluded in hypothesis tests, the sample size of $50 > 42$ was adequate to reject null hypothesis with power = 0.9 at alpha = 0.05 when small effect size of η_p^2 was greater than 0.1.

Materials

Visual stimuli was identical to Experiment 1A, and auditory stimuli was identical to Experiment 1B; visual and auditory sequences were produced by the same grammar as Experiment 1 and were presented simultaneously with each other. An example of cross-modal sequence is presented in Figure 3.

Procedure

In the learning phase, auditory and visual sequences were presented simultaneously, with the visual training set instantiated as animal pictures and audio training set as animal names. The auditory and visual sequences were identical and presented together during the familiarization phase (see Figure 3). For example, every time participants were presented with tiger in the visual modality, they also heard tiger in the auditory modality. At the beginning of the acquisition phase, participants were told nearly the identical story as provided in Experiments 1 and 2, with the additional mention that the animal pictures and names would be presented at the same time.

In the test phase, even though they were trained on a corresponding audiovisual sequence, participants were only tested on the pictures or on the sounds, as was the case in Experiment 1. Participants were randomly assigned to two groups, one group was tested on the visual stimuli, and the other group was tested on the auditory stimuli. The testing phase instructions were identical to the instructions used in Experiments 1 and 2.

Results and Discussion

The mean test accuracy of the visual group in Experiment 2 was 126 out of 160 (79%), with a standard deviation of 21. A one-sample *t*-test indicated that performance on the visual

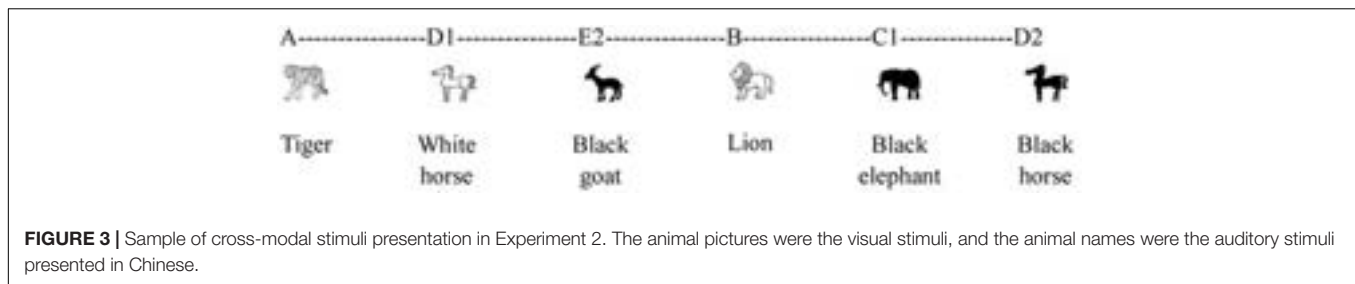


FIGURE 3 | Sample of cross-modal stimuli presentation in Experiment 2. The animal pictures were the visual stimuli, and the animal names were the auditory stimuli presented in Chinese.

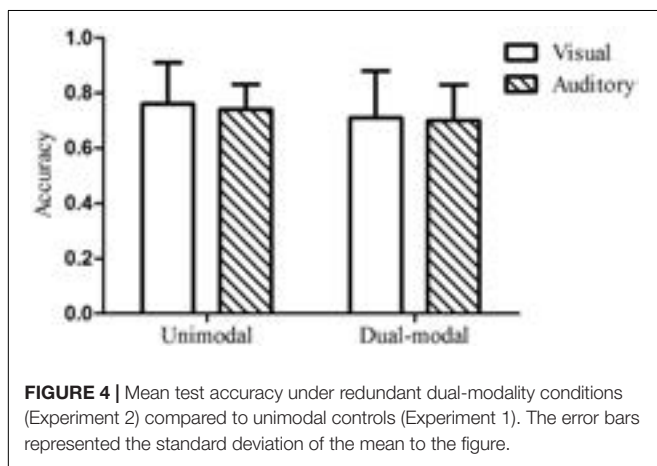


FIGURE 4 | Mean test accuracy under redundant dual-modality conditions (Experiment 2) compared to unimodal controls (Experiment 1). The error bars represented the standard deviation of the mean to the figure.

test was significantly above chance, $t(24) = 11.27$, $p < 0.001$, $d = 2.23$. The mean test accuracy of the auditory group was 115 out of 160 (72%), with a standard deviation of 25. A one-sample t -test indicated that performance on the auditory test was also significantly above chance, $t(24) = 6.86$, $p < 0.001$, $d = 1.38$. A two-way analysis of variance (ANOVA) with condition (dual-modal vs. unimodal baseline) and modality (visual vs. auditory) as between-participants factors was performed. There was no main effect of condition [$F(1,92) = 2.16$, $p = 0.15$, $\eta_p^2 = 0.023$] or modality [$F(1,92) = 0.04$, $p = 0.85$, $\eta_p^2 < 0.001$], nor was there a significant interaction [$F(1,92) = 0.34$, $p = 0.56$, $\eta_p^2 = 0.004$]. These results indicated that there was no enhanced effect of ISL when audiovisual sequences with the same grammatical regularities were presented simultaneously (see **Figure 4**). The very small effect sizes indicate a substantial overlap in the score distributions across the two conditions. To reflect the extent to which the distributions separated across conditions with the effect of this size ($\eta_p^2 = 0.023$), we utilized a priori power analysis based on G*Power version 3.1.9.2 package and found that 2088 participants in total were needed for an effect of this size to reach statistical significance. We further calculated the Bayes Factor (BF) using the R package Bayes Factor (Morey and Rouder, 2015), supplementing the ANOVA, to compare the fit of the data under the null hypothesis and the alternative hypothesis (see von Koss Torkildsen et al., 2018 for a recent example of taking such an approach). The analyses indicated that the data are more in line with the null hypothesis (no difference between conditions), with a BF of 1.80 for the comparison of the null hypothesis over the alternative hypothesis (a difference between conditions) in

contrast to a BF of 0.56 for the comparison of the alternative hypothesis over the null hypothesis, suggesting that the implicit learning performance under dual-modal condition are not differ from that under unimodal condition of either visual or auditory.

The test performance under redundant dual-modality conditions was identical to that under unimodal conditions, which suggests that the underlying learning systems operated in parallel and independently of one another. However, the strongest test of whether dual-modality ISL occurs independently is to present audio-visual sequences simultaneously but with each perceptual modality containing different grammatical regularities.

EXPERIMENT 3: DUAL-MODALITY, CROSS-MODAL ISL WITH DIFFERENT GRAMMARS

The primary goal of Experiment 3 was to examine whether ISL would be decreased when participants were presented with auditory and visual sequences simultaneously with each modality containing different grammatical regularities. If interference is observed, this would suggest that a single ISL system was operating over the cross-modal input, attempting to integrate the sequences into a coherent representation. On the other hand, if no decrement in performance is observed, this would provide strong evidence that the two sets of sequential regularities were processed and learned independently of one another.

Method

Subjects

Forty-nine Chinese graduate students were recruited for the tests under dual-modality condition (Age range = 23–30; Mean age = 25.40; Females = 24). We excluded 1 participant from analysis who failed to follow instructions. An additional 24 Chinese graduate students were recruited to serve as unimodal baseline controls using the new auditory grammar that was used in Experiment 3 (Age range = 24–29; Mean age = 25.88; Females = 13). The power analysis, same as it used in Experiment 2, showed the sample size of 49 > 42 was well-advised to detect a small effect size of $\eta_p^2 = 0.1$ with power = 0.9 at $\alpha = 0.05$.

Materials

The materials were similar to the previous experiments, except that the auditory sequences were generated from a second grammar which was different from the grammar used for the

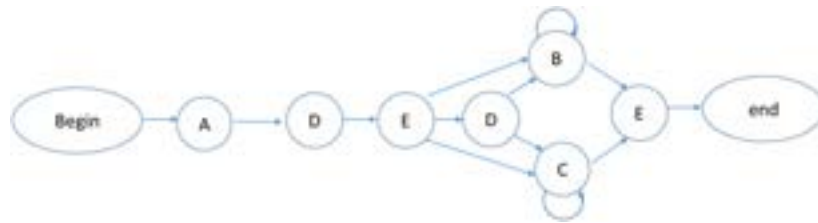


FIGURE 5 | Artificial Grammar 2.

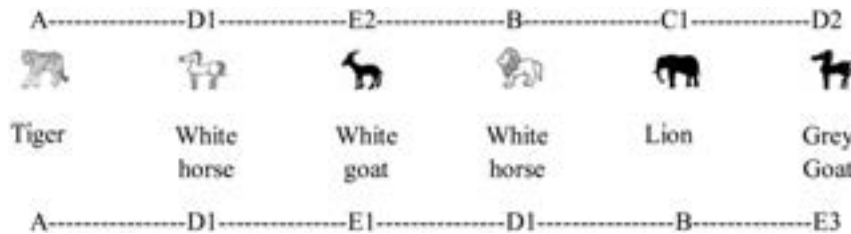


FIGURE 6 | Sample of cross-modal stimuli presentation in Experiment 3. The animal pictures were the visual stimuli (generated from artificial grammar 1) and the animal names were the auditory stimuli (generated from artificial grammar 2) presented in Chinese.

visual sequences. The new grammar is presented in **Figure 5** and the new auditory sequences used in the acquisition and test phases are shown in Appendix A (Grammar 2) and Appendix C, respectively.

Procedure

The procedure was identical to Experiment 2, except for the different auditory sequences. For example, if participants were presented with tiger in the visual modality, it was very likely they heard the name of a different animal (such as lion) in the auditory modality because unlike Experiment 2, there was no cross-modal redundancy. **Figure 6** shows an example of cross-modal sequence used in Experiment 3. In the test phase, even though they were trained on cross-modal sequences, participants were only tested on the visual or auditory sequences, as was the case in Experiment 1. Participants were randomly assigned to two groups, one group was tested on the visual stimuli, and the other group was tested on the auditory stimuli. Instructions for the learning and test phases were identical to Experiment 2.

Because we used a new grammar for the auditory sequences in Experiment 3. It is necessary to obtain a new unimodal baseline. Thus we also collected the ISL performance of the new grammar under auditory unimodal conditions.

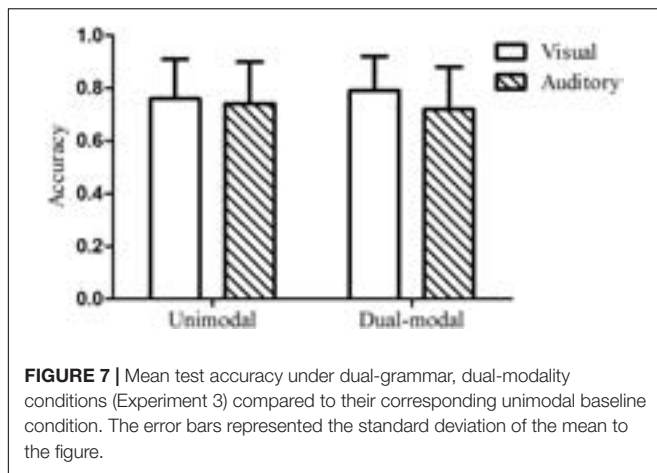
Results and Discussion

The mean test accuracy of the visual group in Experiment 3 was 114 out of 160 (71%), with a standard deviation of 27. A one-sample t -test indicated that performance on the visual test was significantly above chance, $t(23) = 6.12$, $p < 0.001$, $d = 1.24$. The mean test accuracy of the auditory group was 112 out of 160 (70%), with a standard deviation of 21. A one-sample t -test indicated that performance on the auditory test was also significantly above chance, $t(23) = 7.70$, $p < 0.001$,

$d = 1.54$. The mean test accuracy of the new grammar under auditory unimodal conditions was 118 out of 160 (74%), with a standard deviation of 14. A two-way analysis of variance (ANOVA) with condition (dual-modal vs. unimodal baseline) and modality (visual vs. auditory) as between-participants factors was performed. There was no main effect of condition [$F(1,90) = 0.15$, $p = 0.70$, $\eta_p^2 = 0.002$] or modality [$F(1,90) = 3.61$, $p = 0.06$, $\eta_p^2 = 0.039$], nor was there a significant interaction [$F(1,90) = 0.02$, $p = 0.90$, $\eta_p^2 < 0.001$]. The new auditory unimodal baseline from this experiment was used. These results indicated that there was no interference when audiovisual sequences with different grammatical regularities were presented simultaneously (see **Figure 7**). A power analysis showed that we would need 23694 participants in total for an effect of this size ($\eta_p^2 = 0.002$) to reach statistical significance. A Bayesian analysis also indicated that the data were more consistent with the null hypothesis of no difference between conditions, as indicated by a BF of 4.31, as opposed to a BF of 0.23 for the alternative hypothesis of a difference between conditions.

A one-way ANOVA (Experiment 1A visual baseline, Experiment 2 visual modality, and Experiment 3 visual modality) revealed no significant difference in visual test scores, $F(2,68) = 1.97$, $p = 0.15$, $\eta_p^2 = 0.055$. A power analysis showed that we would need 222 participants in total for an effect of this size to reach statistical significance. The BF analysis suggested that the data are more in line with the null hypothesis of no condition effect, as indicated by a BF of 1.74 for the comparison of the null hypothesis over the alternative hypothesis in contrast to a BF of 0.57 for the comparison of the alternative hypothesis over the null hypothesis.

A second one-way ANOVA (supplementing auditory baseline, Experiment 2 auditory modality, and Experiment 3 auditory modality) revealed no significant difference in auditory test scores



across all experiments, $F(2,70) = 0.85$, $p = 0.43$, $\eta_p^2 = 0.024$. Learning under dual-stream conditions in both Experiment 2 (same grammar) and Experiment 3 (different grammars) had essentially the same levels of learning as under unimodal conditions, indicating that the cross-modal presentation format did not influence learning of each perceptual input stream. A power analysis showed that we would need 519 participants in each condition for an effect of this size to reach statistical significance. The BF analysis suggested that the data are more in line with the null hypothesis of no condition effect, as indicated by a BF of 2.84 for the comparison of the null hypothesis over the alternative hypothesis in contrast to a BF of 0.35 for the comparison of the alternative hypothesis over the null hypothesis. The fact that participants were able to simultaneously learn two different sets of sequential regularities from two different perceptual modalities is consistent with the involvement of independent and separate learning subsystems.

GENERAL DISCUSSION

The primary goal of this research was to test systematically whether dual-modality input streams would affect learning for each separate modality. In Experiment 1, we presented visual and auditory streams in isolation to provide baseline levels of unimodal performance used as comparisons for Experiments 2 and 3. In Experiment 2, we presented learners simultaneously with cross-modal streams that consisted of redundant information, that is, the identical sequences using the same artificial grammar. We found no effect on learning under such dual-stream conditions; participants were able to learn both visual and auditory streams successfully and the ISL effect was not enhanced relative to Experiment 1, the baseline level.

The nervous systems of many organisms have the ability to integrate perceptual inputs across different sensory modalities (e.g., visual and auditory), and then to determine the relationship between the corresponding inputs. This generally results in an enhancement for processing of the cross-modal information, allowing the organism to respond more accurately and quickly when processing multiple sources of sensory information from

the same object at the same time (Liu et al., 2010). That there was no facilitative effect observed under redundant cross-modal conditions in Experiment 2 (using the pictures of animals paired with the auditory name of each) suggests that the sequential regularities were processed and learned independently within each perceptual modality, with no cross-modal integration occurring.

A complementary way to examine the effect of cross-modal input on ISL is to present learners with audiovisual streams containing different sequential regularities in each modality. This was explored in Experiment 3, in which we examined whether ISL would be decreased due to the interference between the two modalities. Incredibly, the results showed that this manipulation resulted in no decrement to performance relative to baseline learning conditions. That participants demonstrated learning of the sequential regularities for each input modality, despite each input stream providing different and conflicting information, is perhaps the strongest evidence yet that ISL consists of separate perceptual learning mechanisms operating in parallel (Conway and Christiansen, 2006; Frost et al., 2015).

If ISL is based on a common central mechanism, then dual-modality input conditions should result in cross-modal effects (either facilitation or interference, depending on the nature of the cross-modal relations). This was not observed. There was no difference in the levels of learning for either modality across any of the familiarization conditions (unimodal vs. cross-modal), suggesting that learning in each modality proceeded independently and did not impact each other. The existence of independent learning mechanisms is consistent with the findings of Jiménez and Vázquez (2011) who assessed whether sequence learning and contextual cueing learning could occur simultaneously. They found that it is possible to learn simultaneously about both context information and sequence information (consisting of a relatively complex sequence of targets and responses), thus suggesting that these two learning processes do not compete for a limited pool of central cognitive resources. Their findings provided support for the existence of multiple parallel learning subsystem. A complementary finding was recently observed by Walk and Conway (2016). Their study demonstrated that when visual and auditory stimuli were interleaved in a sequential presentation format, adults readily learned dependencies between stimuli within the same modality (visual–visual or auditory–auditory) but were unable to learn cross-modal sequential dependencies (auditory–visual or visual–auditory).

In contrast, a growing body of evidence has demonstrated cross-modal effects during ISL. For instance, Cunillera et al. (2010) showed that simultaneous visual information could improve auditory ISL if the visual cues were presented near transition boundaries (see also Sell and Kaschak, 2009; Mitchel and Weiss, 2010; Thiessen, 2010). Similarly, Mitchel et al. (2014) used the McGurk illusion to demonstrate that learners can integrate auditory and visual input during a statistical learning task. These studies seem to provide evidence counter to the lack of behavioral facilitation that we observed in Experiment 2. On the flip side, Mitchel and Weiss (2011) found that when triplet boundaries in visual and auditory input streams were

desynchronized, learning was disrupted. This would appear to provide evidence counter to our Experiment 3, which revealed a lack of interference when the sequential regularities were decoupled.

It is possible that our experimental design was insufficient to detect any differences in learning across the different conditions. However, another reason for the differences in effects across studies is that ours used an artificial grammar, which involves learning and generalization of patterns to novel stimuli, unlike the paradigms used by the other studies reviewed above, which used the traditional triplet task, which does not involve generalization to new sequences but rather requires recognizing familiar triplets from unfamiliar ones. It is possible that learning in these simpler designs that do not involve generalization to new stimuli, reflects more explicit learning and memory processes (relying on the medial temporal lobe, Schapiro et al., 2014) whereas AGL also crucially involves striatal and frontal systems such as basal ganglia and prefrontal cortex (Conway and Pisoni, 2008), which are thought to comprise the procedural memory system (Ullman, 2004). Thus, future research needs to disentangle the role of task in obtaining some of these disparate cross-modal effects.

We should also point out that other AGL tasks, not dissimilar to the one used here, have demonstrated that learning can be disrupted under certain conditions. For instance, Hendricks et al. (2013) showed that when participants engaged in a working memory task concurrently with a standard AGL task, learning was disrupted if the working memory task occurred during the test phase or if the grammar learning task involved transfer to a new stimulus set. Similarly, in Experiment 3 of the Conway and Christiansen (2006) study, learning was disrupted when the stimuli in the two sets of grammars were perceptually similar. These findings show that in principle, the AGL task is sensitive to disruption under specific conditions; thus the lack of such interference effects observed here appears to be meaningful.

In summary, our study revealed, quite remarkably, that participants were just as adept at learning statistical regularities from two streams as from one. Importantly, participants were

able to track simultaneously two sets of sequential regularities regardless of the cross-modal stimulus relationship (i.e., whether the two input streams were coupled or decoupled). These findings add to a growing set of findings that suggest that ISL consists of the involvement of multiple learning subsystems that can operate independently and in parallel. This work further suggests that under certain cross-modal input conditions, information is not integrated across perceptual modalities but instead remains distinct, allowing for a powerful set of resources to encode and represent the perceptual regularities that exist in the environment.

AUTHOR CONTRIBUTIONS

XL, XZ, and WS all contributed to the conception and design of this work and interpretation of data; the drafting and revising of the manuscript; and final approval of the version to be published. XL furthermore carried out the acquisition and analysis of the data. CC contributed to the interpretation of data; drafting and revising of the manuscript; and final approval of the version to be published. YL contributed to the interpretation of data; drafting and revising of the manuscript; and final approval of the version to be published during the response to the reviewers.

FUNDING

This research has been supported by grants from the National Institute on Deafness and Other Communication Disorders (R01DC012037 awarded to CC) and from the National Natural Science Foundation of China (31160201 awarded to WS).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00146/full#supplementary-material>

REFERENCES

- Altmann, G. T. M., Dienes, Z., and Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 899–912. doi: 10.1037/0278-7393.21.4.899
- Aslin, R. N., and Newport, E. L. (2009). “What statistical learning can and can’t tell us about language acquisition,” in *Infant Pathways to Language: Methods, Models, and Research Disorders*, eds J. Colombo, P. McCardle, and L. Freund (New York, NY: Erlbaum), 15–29
- Chang, G. Y., and Knowlton, B. J. (2004). Visual feature learning in artificial grammar classification. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 714–722. doi: 10.1037/0278-7393.30.3.714
- Christiansen, M. H., Conway, C. M., and Onnis, L. (2012). Similar neural correlates for language and sequential learning: evidence from event-related brain potentials. *Lang. Cogn. Process.* 27, 231–256. doi: 10.1080/01690965.2011.606666
- Cleeremans, A., and Jiménez, L. (2002). “CHAPTER ONE Implicit learning and consciousness: a graded, dynamic perspective,” in *Implicit Learning and Consciousness: An Empirical, Philosophical and Computational Consensus in the Making*, eds R. M. French, and A. Cleeremans (London: Psychology Press), 1–45.
- Conway, C. M., and Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 24–39. doi: 10.1037/0278-7393.31.1.24
- Conway, C. M., and Christiansen, M. H. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychol. Sci.* 17, 905–912. doi: 10.1111/j.1467-9280.2006.01801.x
- Conway, C. M., and Christiansen, M. H. (2009). Seeing and hearing in space and time: effects of modality and presentation rate on implicit statistical learning. *Eur. J. Cogn. Psychol.* 21, 561–580. doi: 10.1080/09541440802097951
- Conway, C. M., and Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Ann. N. Y. Acad. Sci.* 1145, 113–131. doi: 10.1196/annals.1416.009
- Cunillera, T., Càmarà, E., Laine, M., and Rodríguez-Fornells, A. (2010). Speech segmentation is facilitated by visual cues. *Q. J. Exp. Psychol.* 64, 1021–1040. doi: 10.1080/17470210902888809

- Durrant, S. J., Cairney, S. A., and Lewis, P. A. (2013). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cereb. Cortex* 23, 2467–2478. doi: 10.1093/cercor/bhs244
- Emberson, L. L., Conway, C. M., and Christiansen, M. H. (2011). Timing is everything: changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Q. J. Exp. Psychol.* 64, 1021–1040. doi: 10.1080/17470218.2010.538972
- Emberson, L. L., and Rubinstein, D. Y. (2016). Statistical learning is constrained to less abstract patterns in complex sensory input (but not the least). *Cognition* 153, 63–78. doi: 10.1016/j.cognition.2016.04.010
- Fiser, J., and Aslin, R. N. (2002). Statistical learning of higher order temporal structure from visual shape-sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 458–467. doi: 10.1037/0278-7393.28.3.458
- Frost, R., Armstrong, B. C., Siegelman, N., and Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn. Sci.* 19, 117–125. doi: 10.1016/j.tics.2014.12.010
- Fu, Q. F., and Fu, X. L. (2006). Relationship between representation and consciousness in implicit learning. *Adv. Psychol. Sci.* 14, 18–22.
- Goujon, A., and Fagot, J. (2013). Learning of spatial statistics in nonhuman primates: contextual cueing in baboons (*Papio papio*). *Behav. Brain Res.* 247, 101–109.
- Guo, X., Jiang, S., Wang, H., Zhu, L., Tang, J., Dienes, Z., et al. (2013). Unconsciously learning task-irrelevant perceptual sequences. *Conscious. Cogn.* 22, 203–211. doi: 10.1016/j.cog.2013.03.004
- Guo, X. Y., Jiang, S., Ling, X. L., Zhu, L., and Tang, J. H. (2011). Specific contribution of intuition to implicit learning superiority. *Acta Psychol. Sin.* 43, 977–982. doi: 10.1016/j.concog.2012.12.001
- Hendricks, M. A., Conway, C. M., and Kellogg, R. T. (2013). Using dual-task methodology to dissociate automatic from non-automatic processes involved in artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 1491–1500. doi: 10.1037/a0032974
- Jiménez, L., and Vázquez, G. A. (2011). Implicit sequence learning and contextual cueing do not compete for central cognitive resources. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 222–235. doi: 10.1037/a0020378
- Johansson, T. (2009). Strengthening the case for stimulus-specificity in artificial grammar learning. *Exp. Psychol.* 56, 188–197. doi: 10.1027/1618-3169.56.3.188
- Kemény, F., and Lukács, A. (2011). Perceptual effect on motor learning in the serial reaction-time task. *J. Gen. Psychol.* 138, 110–126. doi: 10.1080/00221309.2010.542509
- Li, X. J., and Shi, W. D. (2016). Influence of selective attention on implicit learning with auditory stimuli. *Acta Psychol. Sin.* 48, 221–229. doi: 10.3724/SP.J.1041.2016.00221
- Liu, Q., Hu, Z. H., Zhao, G., Tao, W. D., Zhang, Q. L., Sun, H. J., et al. (2010). The prior knowledge of the reliability of sensory cues affects the multisensory integration in the early perceptual processing stage. *Acta Psychol. Sin.* 42, 227–234. doi: 10.3724/SP.J.1041.2010.00227
- Mastroberardino, S., Santangelo, V., Botta, F., Marucci, F. S., and Belardinelli, M. O. (2008). How the bimodal format of presentation affects working memory: an overview. *Cogn. Process.* 9, 69–76. doi: 10.1007/s10339-007-0195-6
- Mitchel, A. D., Christiansen, M. H., and Weiss, D. J. (2014). Multimodal integration in statistical learning: evidence from the McGurk illusion. *Front. Psychol.* 5:407. doi: 10.3389/fpsyg.2014.00407
- Mitchel, A. D., and Weiss, D. J. (2010). What's in a face? Visual contributions to speech segmentation. *Lang. Cogn. Process.* 25, 456–482. doi: 10.1080/01690960903209888
- Mitchel, A. D., and Weiss, D. J. (2011). Learning across senses: cross-modal effects in multisensory statistical learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1081–1091. doi: 10.1037/a0023700
- Morey, R. D., and Rouder, J. N. (2015). *Computation of Bayes Factors for Common Designs. R Package Version 0.9.12-2*. Available at: <http://ppl.missouri.edu/bayesfactor>.
- Peña, M., Bonatti, L. L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science* 298, 604–607. doi: 10.1126/science.1072901
- Reber, A. S. (1967). Implicit learning of artificial grammars. *J. Verbal Learn. Verbal Behav.* 6, 855–863. doi: 10.1016/S0022-5371(67)80149-X
- Reber, A. S. (1989). Implicit learning and tacit knowledge. *J. Exp. Psychol. Gen.* 118, 219–235. doi: 10.1037/0096-3445.118.3.219
- Redington, M., and Chater, N. (1996). Transfer in artificial grammar learning: a reevaluation. *J. Exp. Psychol. Gen.* 125, 123–138. doi: 10.1037/0096-3445.125.2.123
- Rosas, R., Ceric, F., Tenorio, M., Mourgues, C., Thibaut, C., Hurtado, E., et al. (2010). ADHD children outperform normal children in an artificial grammar implicit learning task: ERP and RT evidence. *Conscious. Cogn.* 19, 341–351. doi: 10.1016/j.concog.2009.09.006
- Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52. doi: 10.1016/S0010-0277(98)00075-4
- Santangelo, V., Fagioli, S., and Macaluso, E. (2010). The costs of monitoring simultaneously two sensory modalities decrease when dividing attention in space. *NeuroImage* 49, 2717–2727. doi: 10.1016/j.neuroimage.2009.10.061
- Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., and Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *J. Cogn. Neurosci.* 26, 1736–1747. doi: 10.1162/jocn_a_00578
- Sell, A. J., and Kaschak, M. P. (2009). Does visual speech information affect word segmentation? *Mem. Cogn.* 37, 889–894. doi: 10.3758/MC.37.6.889
- Shanks, D. R., Johnstone, T., and Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Q. J. Exp. Psychol.* 50, 216–252. doi: 10.1080/713755680
- Shi, W. D., Li, X. J., Wang, W., and Yan, W. H. (2013). Comparison of implicit learning effect between multisensory and unisensory. *Acta Psychol. Sin.* 45, 1313–1323. doi: 10.3724/SP.J.1041.2013.01313
- Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2331
- Thiessen, E., and Erickson, L. (2015). Perceptual development and statistical learning. *Handb. Lang. Emerg.* 87, 396–414. doi: 10.1002/978111834613_6.ch18
- Thiessen, E. D. (2010). Effects of visual information on adults' and infants' auditory statistical learning. *Cogn. Sci.* 34, 1093–1106. doi: 10.1111/j.1551-6709.2010.01118.x
- Tunney, R. J., and Altmann, G. T. M. (2001). Two modes of transfer in artificial grammar learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 614–639. doi: 10.1037/0278-7393.27.3.614
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative procedural model. *Cognition* 92, 231–270. doi: 10.1016/j.cognition.2003.10.008
- von Koss Torkildsen, J., Arciuli, J., Haukedal, C. L., and Wie, O. B. (2018). Does a lack of auditory experience affect sequential learning? *Cognition* 170, 123–129. doi: 10.1016/j.cognition.2017.09.017
- Walk, A. M., and Conway, C. M. (2016). Cross-modal statistical-sequential dependencies are difficult to learn. *Front. Psychol.* 7:250. doi: 10.3389/fpsyg.2016.00250

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Li, Zhao, Shi, Lu and Conway. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.