Influence of Perceptual Saliency Hierarchy on Learning of Language Structures: An Artificial Language Learning Experiment

Tao Gong[1,2]*, Yau W. Lam[3] and Lan Shuai[1]

[1] Haskins Laboratories, New Haven, CT, USA, [2] Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China, [3] Department of Linguistics, University of Hong Kong, Hong Kong, China

Psychological experiments have revealed that in normal visual perception of humans, color cues are more salient than shape cues, which are more salient than textural patterns. We carried out an artificial language learning experiment to study whether such perceptual saliency hierarchy (color > shape > texture) influences the learning of orders regulating adjectives of involved visual features in a manner either congruent (expressing a salient feature in a salient part of the form) or incongruent (expressing a salient feature in a less salient part of the form) with that hierarchy. Results showed that within a few rounds of learning participants could learn the compositional segments encoding the visual features and the order between them, generalize the learned knowledge to unseen instances with the same or different orders, and show learning biases for orders that are congruent with the perceptual saliency hierarchy. Although the learning performances for both the biased and unbiased orders became similar given more learning trials, our study confirms that this type of individual perceptual constraint could contribute to the structural configuration of language, and points out that such constraint, as well as other factors, could collectively affect the structural diversity in languages.

Keywords: perceptual saliency hierarchy, artificial language learning, syntax, learning bias, diversity

INTRODUCTION

Physical objects can be discriminated by visual features such as color, shape, and texture. Human eyes are essentially light receptors, and thus, color or brightness information requires little cognitive load for processing, thus becoming the strongest cue for visual perception. In terms of evolution, the alimentary "niche" also enhanced color perception in humans and other primates (Dominy and Lucas, 2001; Melin et al., 2012). Difference in color or brightness enables humans to perceive additional features such as shape and textural pattern. Per these fundamental features (color, shape, and texture), psychological experiments have explicitly shown that: random variations in color interfere with viewer's ability to identify shapes, but variations in shape have no explicit effects (in terms of judgement accuracies and reaction times) on color discrimination (Callaghan, 1990; Healey, 2000); and random variations in color or shape interfere with viewer's identification of visual patterns of texture, but not vice-versa (Treisman, 1985; Healey and Enns, 1999). This evidence reveals a *perceptual saliency hierarchy* (PSH, the relative conspicuousness of various visual features at first exposure, Healey and Enns, 2012), which states that in normal visual

perception of humans, color information appears to be more salient than shape information, and shape more salient than visual textural pattern (simply, color > shape > texture).

Language serves as the primary means for humans to describe visual features. Given the PSH, an interesting question arises: Whether the PSH can cast any influence on learning or processing the language structures used to regulate the relevant adjectives of those visual features. Answer to this question helps reveal the relationship between structural configuration in language and perceptual or cognitive constraints in humans, which is a challenging issue in modern psychology and linguistics (Christiansen and Kirby, 2003; Gentner and Goldin-Meadow, 2003; Hurford, 2007, 2012).

Many approaches have been adopted to study this issue. Corpus analyses have identified universal characteristics in language structures and potential links between language structures and cognitive constraints in humans (Ferrer-i-Cancho, 2004; Liu, 2010; Futrell et al., 2015). Computational modeling (Gong et al., 2013, 2014) has demonstrated how psychological or physiological constraints help shape word order (Gong et al., 2009), compositionality (Kirby, 1999; Brighton, 2002; Smith et al., 2003; Kirby et al., 2006), and syntactic patterns such as recursion, case, or long-distance dependency (Elman, 1990; Conway and Christiansen, 2001; Christiansen and Ellefson, 2002; Lupyan and Christiansen, 2002; Reali and Christiansen, 2009; Christiansen and Chater, 2015). In particular, some simulations show that the word order bias (in favor of certain orders like SOV or SVO but against others like VOS or OVS) in the world's languages could result from individual perceptual constraint, which takes effect during communications (Gong et al., 2009). Other simulations illustrate that the universal color naming patterns in the world's languages could result from the perceptual constraint of human eyes towards colors, which also takes effect during cultural transmission of color terms (Baronchelli et al., 2012). These studies have illustrated the effect of perceptual or cognitive constraints on structural configuration of language (Heine and Kuteva, 2008; Chater et al., 2009; Mesoudi, 2011; Richerson and Christiansen, 2013).

In experimental psychology, the paradigm of artificial language learning (ALL, in which participants are asked to learn a language or language-like system, and then tested on what they have learned; depending on underlying structures, ALL is also called artificial grammar learning) has been used to investigate issues concerning language and cognition (Esper, 1925; Reber, 1967; Folia et al., 2010; Onnis, 2012). An ALL experiment typically consists of a sequence of learning (a.k.a. training) and testing phases, which alternate throughout the experiment. In a learning phase, participants are presented with visual or auditory symbols concatenated following a predefined grammar-like structure. In the subsequent testing phase, they are presented with already-seen or unseen instances. Individual learning is said to occur when participants can distinguish instances that respect the underlying structure from those that violate it.

Artificial language learning experiments can design pseudo-words and structures distinct from participants' native language to diminish the influence of participants' prior linguistic knowledge and highlight corresponding learning mechanisms

and factors hard to control in naturalistic scenarios (Onnis, 2012). They can also generate sufficient instances to trace individual learning and evaluate whether individuals can generalize their learned knowledge to unseen instances. In addition, by recruiting human participants and carefully designed artificial languages, ALL experiments can complement other approaches, such as verifying simulated behaviors and modeling results to bridge the gap between language processing in humans and relevant mechanisms in artificial agents (Kirby et al., 2008; Cornish, 2010). Furthermore, it has been repetitively shown that ALL experiments can uncover the same (or similar) mechanisms manifest in natural and artificial language processing (Reber, 1993; Gómez and Gerken, 2000; Pothos, 2007) and in first (Misyak et al., 2010; Misyak and Christiansen, 2012) and second language acquisition (Friederici et al., 2002; Robinson, 2010; Brooks et al., 2011; Petersson et al., 2012; Morgan-Short et al., 2014; Ettlinger et al., 2015). These advantages have made ALL experiments revitalize language learning research in the past century (Braine, 1963; Moeser and Bregman, 1972; Saffran et al., 1996; Morgan and Newport, 1981; Tily et al., 2011; Tabullo et al., 2012).

To our knowledge, there are no modeling or experimental studies that address directly the PSH and its influence on language learning. In this paper, we conducted an ALL experiment to study this issue. A number of artificial languages were designed, each describing two out of the three types of visual features in the PSH. In an artificial language, a visual feature was mapped to a phonetic segment, and segments, respectively, encoding the two features followed a consistent order. We referred to the theme-first principle in linguistics to clarify such orders. The principle states that more "thematic" information tends to precede less "thematic" one in normal linguistic expressions (Tomlin, 1986) (here, thematic information refers to the pragmatic or psycholinguistic reflex of the general attention in human cognition). This principle helps account for many cross-language phenomena, especially for word order (Halliday, 1967; Mathesius, 1975; Lambrecht, 1994; Cinque, 1999, 2010; Cinque and Rizzi, 2008; Longobardi and Guardiano, 2009). In terms of visual perception, it suggests that information of perceptually more salient feature should precede that of less salient feature. Following this principle, in our ALL experiment, we regarded an order as *congruent* with the PSH, if it puts a segment encoding a perceptually more salient feature in front of a segment encoding a less salient feature; otherwise, the order is deemed *incongruent*. We recruited human participants to learn, by repetitive exposure of instances, the artificial languages having congruent or incongruent orders, and assessed individual learning using seen and unseen instances.

In the following sections, we described the experiment, reported its results, and discussed the relation between language and human cognition based on this study.

MATERIALS AND METHODS

The experimental protocol was approved by the College Research Ethics Committee of University of Hong Kong. The methods

were carried out in accordance with the approved guidelines from the College Research Ethics Committee. Informed consents were obtained from all participants.

Participants

One hundred and thirty-two students from the University of Hong Kong participated into the experiment (66 females, mean age = 23.27, age range = 17–33, SD = 4.97). Forty Hong Kong dollars were paid to participants who had finished the experiment and filled in the post-experiment surveys. The recruited participants had normal or adjusted-to-normal vision, and reported no history of developmental delay or acquired neurological disorder. They were native Mandarin or Cantonese speakers, and had an intermediate level of English.

Materials

We defined six artificial languages, respectively, used in six experimental conditions (see **Figure 1**). Each artificial language described two of the three visual features in the PSH (color (C) and shape (S), color and textural pattern (T), and shape and textural pattern). There are two reasons for considering orders

between only two visual features. First, as shown in previous psychological experiments of visual feature saliency (Treisman, 1985; Callaghan, 1990; Healey and Enns, 1999; Healey, 2000), using languages encoding only two visual features can directly reflect whether the congruent or incongruent orders between the two affect the learning of those orders. Second, training participants on orders among three visual features would require more learning trials to give participants enough opportunities to detect and learn similarities between visual features and segments and similarities in regulating orders among segments. This would increase learning difficulty and memory burden, extend experiment time, and might have adverse effects on participants' motivation.

Our training stimuli consisted of 48 images created by PhotoImpact X3. Each image depicted an object with a unique combination of a shape (star, square, triangle, or circle), a color (red, yellow, green, or blue), and a textural pattern (stripes, dots, zigzag, or checkerboard). These images were divided evenly into three sets. Each set of 16 images differed in two features (e.g., color and shape) and were the same in the third (e.g., texture) (see **Figure 1**).

FIGURE 1 | Meaning-form mappings of the artificial languages in the six experimental conditions. In each table, the rows and columns list the eight instances of the two types of visual feature (four in each type), and each cell shows the form encoding the stimuli having the features specified by the row and column. Each table shows the 16 meaning-form mappings of an artificial language. Hyphens in forms are added to highlight segments. In the actual experiment, participants are exposed to forms without hyphens or other indicators of structure.

All forms of the artificial languages were presented visually in the experiment. A form consisted of two compositional segments. A segment encoded one instance of a visual feature and had a consonant-vowel or vowel structure. All segments had roughly the same level of learning difficulty, and did not resemble any orthography of real words in English or any pronunciation of real characters in Mandarin or Cantonese. We also designed the segments to avoid iconicity (perceptuomotor analogies between aspects of a form and meaning of a word, e.g., onomatopoeia words and ideophones, Dingemanse, 2012; Dingemanse et al., 2015), which could assist language learning or comprehension (Simner et al., 2010; Pemiss and Vigliocco, 2014). In each form of an artificial language, the two segments followed a consistent order. Depending on encoded visual features, the order between segments was either congruent or incongruent with the PSH.

As shown in **Figure 1**, languages 1 and 2 describe color (C) and shape (S), languages 3 and 4 described color (C) and texture (T), and languages 5 and 6 described S and T. Each pair of the languages were formed by the same set of segments but differed in regulating order. Three of these languages (languages 1, 3, and 5) had congruent orders (CS, CT, and ST), and the other three (languages 2, 4, and 6) had incongruent orders (SC, TC, and TS).

Procedure

The procedure was implemented using E-Prime 2.0. During the experiment, participants sat comfortably in front of a laptop in a bright, quiet room. They were asked to learn an "alien language" by viewing its meaning-form mappings displayed on a 21-inch computer monitor at a resolution of 1280 × 1024 and a refresh rate of 75 Hz. The font size was 64 pixels. The distance between the screen and participants' eyes was approximately 64 cm. We used a between-subject design; each participant was assigned to one experimental condition to learn the corresponding artificial language. Gender and number of participants were balanced in each condition (11 females and 11 males). Prior to the experiment, the participants went through a two-minute familiarization block.

The experiment consisted of three 5-min blocks, with optional two-minute breaks in between; the whole experiment lasted about 20 min. A block consisted of a learning and a testing phase; in total, there were three learning and three testing phases to trace learning progress.

In a learning phase, 12 out of the 16 meaning (image)-form mappings (those in the white cells in **Figure 1**) of an artificial language were displayed visually to participants. A mapping was shown on the center of the monitor, with the form presented simultaneously underneath the image (see **Figure 2A**). A mapping remained visible for five seconds. Presentation of all 12 mappings was repeated three times. Each time the meaning-form pairs were displayed in a pseudo-random order ensuring that the images of any two consecutively presented mappings shared no instances of the two types of visual features that the artificial language described. This setting prevented the participants from immediately noticing the associations between the visual features and the segments, thus increasing the difficulty of the learning task.

In a testing phase, individual learning was assessed by 20 forced-choice questions presented in a pseudo-random order. Participants gave their answers by key pressing. After the participants answered a question, the next one popped up without feedback. Ten of the questions were meaning selection questions (see **Figure 2B** for an example). In each of them, the participants saw a form followed by three meanings (images) displayed in a pseudo-random order. Participants were asked to select the image that they believed was expressed by the form. Incorrect meanings shared at most one instance of the visual feature with the correct one. The other ten questions were form selection questions (see **Figure 2C** for an example). In each of them, one image and three forms were displayed simultaneously. The participants were asked to select the form that they believed encoded the meaning. Incorrect forms shared at most one segment with the correct form. The segment orders in the incorrect forms were distinct from the order used in the instances in the learning phase.

The 12 meaning-form mappings shown in the learning phases appeared at least once and at most twice as the correct answers in the 20 testing questions. Each mapping had the same occurrence frequency in the learning and testing phases. To answer the testing questions correctly, the participants needed to learn not only the mappings between the visual features and the segments but also the order between the segments. Compared with the much larger search space in the free recall tasks as in previous studies (e.g., Cornish, 2010; Tamariz and Kirby, 2015), answers in the forced choice questions were more
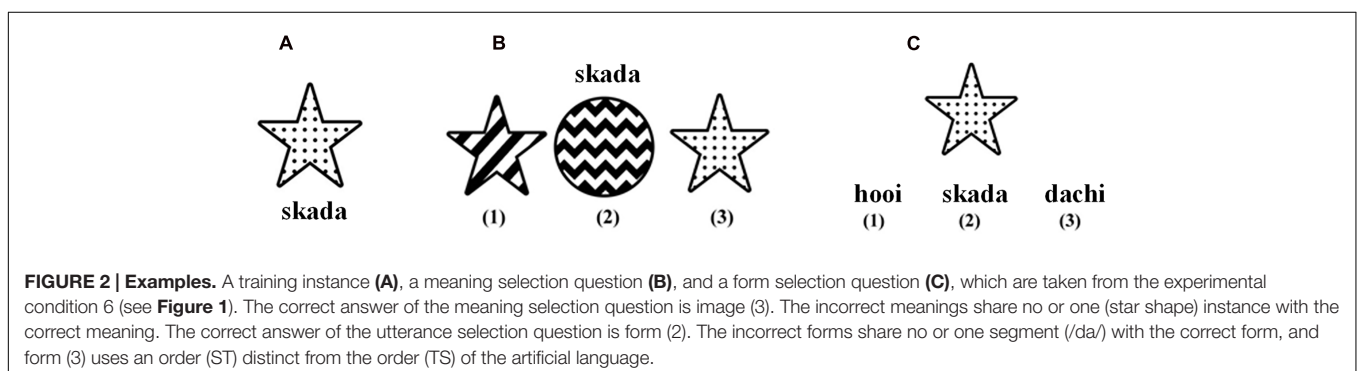
FIGURE 2 | Examples. A training instance **(A)**, a meaning selection question **(B)**, and a form selection question **(C)**, which are taken from the experimental condition 6 (see **Figure 1**). The correct answer of the meaning selection question is image (3). The incorrect meanings share no or one (star shape) instance with the correct meaning. The correct answer of the utterance selection question is form (2). The incorrect forms share no or one segment (/da/) with the correct form, and form (3) uses an order (ST) distinct from the order (TS) of the artificial language.

limited and allowed explicitly tracing the participants' learning performances.

In the last testing phase, apart from the 20 normal testing questions containing the items already seen in the learning phases, there were additional four meaning selection and four form selection questions that contained the novel meaning-form mappings not presented in the learning phases (those in the gray cells in **Figure 1**). Performance on these items helped evaluate whether the participants could generalize their learned knowledge to unseen instances. All the 28 questions were presented in a pseudo-random order.

Measures

In each testing phase, we recorded each participant's accuracy (percentage of correct answers to the questions of the same type) and average reaction time to each of the meaning and form selection questions. In the last testing phase, apart from the measures to the normal testing questions, we also recorded the accuracies and average reaction times to the additional questions about the novel items. We grouped the accuracy and average reaction time data in the experimental conditions 1, 3, and 5 (the artificial languages therein had congruent orders) as the congruent set, and those in the conditions 2, 4, and 6 as the incongruent set. In each set, the accuracies and average reaction times were grouped according to the three testing phases. The measures to the additional questions formed the fourth phase. To meet the assumption of normality, we used the log-transformed (base e) reaction times in the analyses.

After the experiment, the participants were asked to fill a *post hoc* survey to indicate: which type of questions – meaning or form selection – was harder to answer; in which block they could confidently learn the "alien language"; and how difficult they felt to learn the "alien language" on a scale of 1 to 5, '1' being the easiest, '3' being neutral, and '5' being the hardest.

Preprocessing and Analyses

Following the general procedure in assessing experimental data (Osborne and Overbay, 2004), we removed the outliers from the accuracy and reaction time data before the analyses. Outliers were values exceeding 2.5 standard deviations from the group mean. For accuracies, outliers were accuracies that were too low; for reaction times, they were times either too long or too short. Among the 1056 ($132 \times 4 \times 2$) accuracy data in eight groups, 34 outliers were removed; for the reaction times, 23 were removed. Another way to handle outliers is to replace them with the group

means, the results following this procedure were similar (see Supplementary Table S1).

We conducted two ANOVAs, respectively, on accuracy and average reaction time to test our working hypothesis that the PSH affects the learning of regulating orders between segments encoding the involved visual features. In the ANOVAs, we treated the congruency of artificial languages as a between-subjects factor (two levels: the congruent languages, those used in conditions 1, 3, and 5, and incongruent languages, those used in conditions 2, 4, and 6), and the experimental phase as a within-subjects factor (four levels: the testing phases 1, 2, 3, and 4, the latter of which consists of the measures to the additional testing questions involving the unseen items). The ANOVA tests also took into account the question type (two levels: meaning selection or utterance selection) and interaction between congruency and experimental phase. In addition to the ANOVA tests, we conducted group t-tests to compare the accuracies and average reaction times between the conditions differing in regulating orders (conditions 1 vs. 2, 3 vs. 4, and 5 vs. 6), which aimed to reveal possible learning biases for the congruent or incongruent orders. Following the Bonferroni correction, we set the critical p value to identify significant effects as 0.002 (0.05/(2+24), 26 tests in total). All the analyses were carried out in R 3.2.4 (R Core Team, 2016).

RESULTS

Table 1 shows the results of the ANOVAs. In both tests, question type showed no significant effect, which matched the *post hoc* surveys; 125 participants felt invariant to both types of questions. This indicated that the way of recording the individual learning performance in our study had no obvious effect on the recorded results.

Both congruency and experimental phase showed significant main effects, but there was no significant interaction between the two. Compared with experimental phase, congruency had a smaller effect size η^2. The significant effect of congruency confirmed our working hypothesis that the perceptual saliency hierarchy could affect individual learning of congruent or incongruent orders. The significant effect of experimental phase indicated that learning occurred at different experimental phases of instance exposure. The non-significant interaction between congruency and experimental phase suggested that the learning patterns across phases for the congruent and incongruent orders were largely the same.

TABLE 1 | Results of the ANOVAs of accuracy (ACC) and average reaction time (RT).

Factor	ANOVA of ACC			ANOVA of RT		
	F	**P**	**η²**	**F**	**p**	**η²**
Congruency	**30.413**	**<0.00005**	**0.029**	**17.442**	**<0.00005**	**0.017**
Phase	**40.006**	**<0.00001**	**0.106**	**24.090**	**<0.00001**	**0.066**
Question Type	0.819	0.366		0.558	0.455	
Congruency × Phase	4.439	0.004		2.867	0.036	

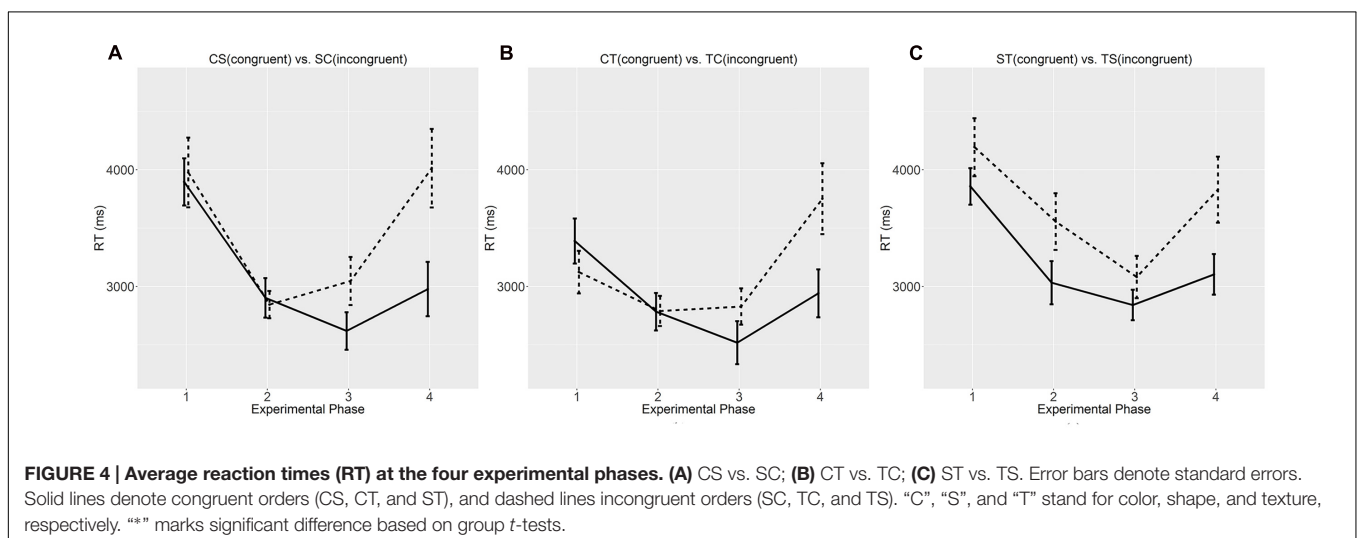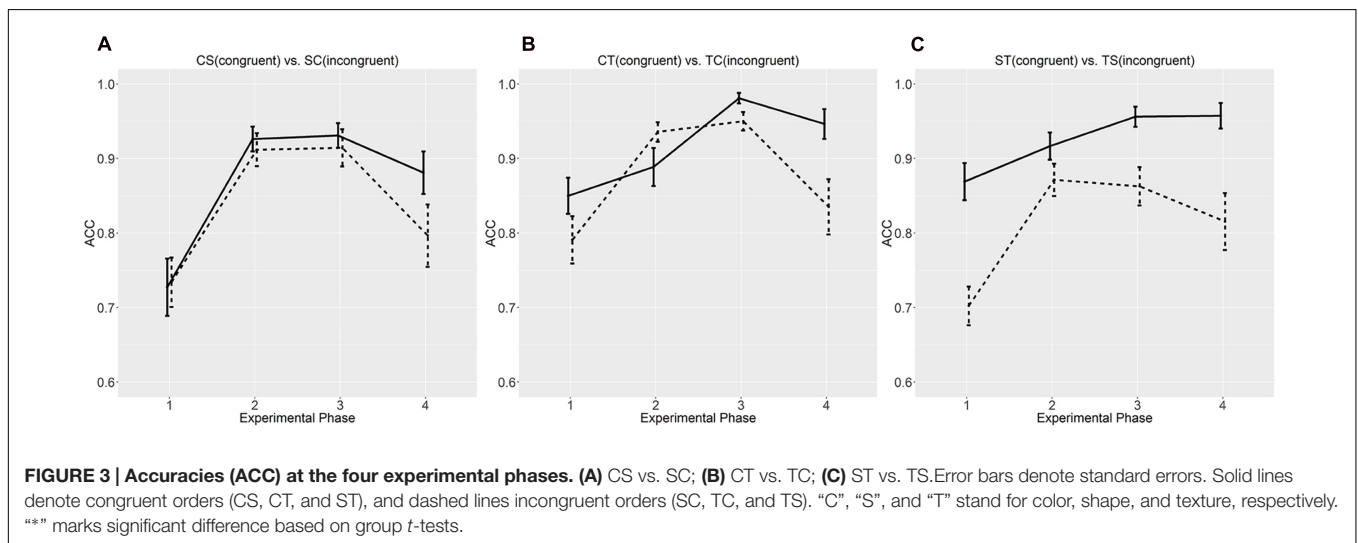*Significant effects (whose p-values are below the critical p-value 0.002) are highlighted in bold.*

Figures 3 and **4** compare the accuracies and average reaction times across the four phases between the conditions differing in regulating orders. Across the three phases, there were a general increase in accuracy and a general decrease in average reaction time, which echoed the significant effect of experimental phase in the ANOVAs. In our experiment, individual learning started at the first phase, and the improvement after the third phase was smaller than that after the second phase. After three phases, the participants had largely grasped the compositional languages in different conditions. These results also matched the *post hoc* surveys; 128 participants claimed that they had confidently learned most of the meaning-form mappings after the second phase, and the other four said that they had learned the language right after the first phase.

Although the participants claimed to have learned the artificial languages, their performances on the unseen items at the last phase revealed some biases for the congruent orders. As for the CS and SC orders, the participants showed similar accuracies, but their reaction times to the CS orders were shorter than those to the SC orders. As for the CT and TC orders, they showed higher accuracies and shorter reaction times to the CT order than the TC order. As for the ST and TS orders, they showed a similar bias for the ST order over the TS order. Significant difference in accuracy was also shown at the first and third testing phases concerning the seen instances. These biases also manifested in the *post hoc* surveys; when asked to evaluate the learning difficulty of the "alien language", the average scores given by the participants in the CS and SC conditions were similar (2.38 vs. 2.64), but those in the CT and ST conditions were different (2.86 vs. 3.65), so were their scores in the ST and TS conditions (3.10 vs. 3.95).

DISCUSSION

In this paper, we evaluated whether the perceptual constraint regarding the saliency hierarchy of the basic visual features affects the learnability of ordering structures between the segments encoding such features in an artificial language. After repeated

FIGURE 3 | Accuracies (ACC) at the four experimental phases. (A) CS vs. SC; (B) CT vs. TC; (C) ST vs. TS. Error bars denote standard errors. Solid lines denote congruent orders (CS, CT, and ST), and dashed lines incongruent orders (SC, TC, and TS). "C", "S", and "T" stand for color, shape, and texture, respectively. "*" marks significant difference based on group *t*-tests.

FIGURE 4 | Average reaction times (RT) at the four experimental phases. (A) CS vs. SC; (B) CT vs. TC; (C) ST vs. TS. Error bars denote standard errors. Solid lines denote congruent orders (CS, CT, and ST), and dashed lines incongruent orders (SC, TC, and TS). "C", "S", and "T" stand for color, shape, and texture, respectively. "*" marks significant difference based on group *t*-tests.

exposure to the tokens of the artificial languages with different orderings, the participants gradually learned the segments encoding color, shape, or textural patterns and the orders between these segments. Their judgements on the unseen instances indicated that they could generalize their learned knowledge and apply it to novel items. Moreover, they exhibited biases for the orders that were congruent with the perceptual saliency hierarchy regarding color, shape, and textural patterns. To be specific, they showed strong biases for the CT (color before textural pattern) and ST (shape before textural pattern) orders over the TC and TS orders, in terms of judging accuracy and average reaction time. Such biases started to exhibit during the learning process. They also showed a weak bias for the CS over the SC order, which only manifested in average reaction time when judging the unseen items.

In this ALL experiment, the observed biases were not induced by participants' prior linguistic knowledge (of Mandarin, Cantonese, or English). In simple phrases of Mandarin or Cantonese, information of textural pattern often appears before that of shape, and color before shape (e.g., "*hongse* (red) *mutou* (wood) *yuan* (round) *zhuozi* (table)") (Yip and Rimmington, 2004; Zhu, 2005), whereas the participants in our experiment exhibited a strong bias for the orders putting textural pattern after color or shape. In simple phrases of English, adjectives of shapes often appear in front of those of colors (e.g., "a round red wood table") (Carter et al., 2011), but the participants showed no bias for color and shape at least in accuracy. In addition, the participants had no previous experience of the segments used in our experiment, and had no chance to apply their prior linguistic knowledge to change the artificial languages or develop one from scratch. These ensure that the observed patterns can be safely ascribed to the perceptual saliency hierarchy. Nonetheless, we acknowledge that participants' alphabetic knowledge may potentially affect their performance in ALL. This is an inevitable limitation of ALL experiments recruiting alphabetic language speakers to learn alphabetic languages. Recruiting participants with no alphabetic experiences (e.g., pre-language children) or using uncommon symbols or non-linguistic forms to design artificial languages may help diminish such influence, as in experimental semiotics studies (Galantucci and Garrod, 2010, 2011) (e.g., Galantucci, 2005; Scott-Phillips et al., 2009; Taylor et al., 2011; De Boer and Verhoef, 2012; Claidière et al., 2014; Tamariz and Kirby, 2015). However, many of such studies focus on the emergence of a language-like communication system out of random signals, and participants therein are allowed to introduce signals that they prefer during the recall tasks.

Our fact that the perceptual saliency hierarchy affects the learning and processing of relevant language structures reveals a close relation between perceptual constraints in humans and structural configuration in language. In a linguistic form, if the ordering of segments encoding the visual features follows naturally the perceptual saliency of those features, production and comprehension of the form would be more straightforward. Then, compared with the forms having an incongruent order between those features, the accuracies of answering questions about the forms having congruent orders tend to be higher, and

the reaction times shorter. This is evident by the strong biases for the CT and ST orders over TC and TS orders.

In addition, compared with color and shape, textural pattern is much less salient, and it is shown as contrast of color or shape. Detection of such pattern occurs after detection of color or shape, and relies on detection of color or shape (Treisman, 1985; Healey and Enns, 1999). This also explains the strong bias for the CT and ST orders. By contrast, color appears to be slightly more salient than shape, which resulted in the weak bias for the CS over SC order.

All these results are in line with the perspective that perceptual constraints affect the learning (and use) of related language structures (Heine, 2001; Pullum and Scholz, 2002; Mesoudi and Whiten, 2008; Christiansen and Chater, 2015). They also suggest that difference in saliency levels of the visual features could affect the degree of bias for the congruent orders regulating those features.

Given the bias for the congruent orders, a follow-up question arises: If the perceptual saliency hierarchy affects the regulating orders of the segments encoding the involved visual features, is this structure in all languages the same, favoring the congruent orders? The answer to this question is NO. Although there lack large-scale typological studies of adjective orders in world's languages, as shown in simple Chinese expressions, the adjectives of textural patterns usually appear in front of the color or shape adjectives. Although in many English phases, the shape adjectives should appear in front of the color adjectives, most people have a relatively free order between the two. Considering these, apart from the perceptual saliency hierarchy, the structural configuration of language is also subject to other constraints. One candidate of such constraints comes from the socio-cultural environment of language. As shown in typological studies of structural diversity in world's languages (Haspelmath, 2007; Evans and Levinson, 2009; Dunn et al., 2011), cultural histories of speakers and contact histories between different languages could induce different types of structure.

In addition, our experiment showed that despite the fact that the participants exhibited biases towards the congruent orders, after a small number of learning rounds, they could largely grasp both the biased and unbiased orders, reaching high (over 0.8) accuracy and short reaction time. Following the dynamics in the three experimental phases, we can reasonably expect that given more rounds of learning the participants would learn each artificial language equally well, no matter whether its order was congruent or incongruent. This suggests that the structures distinct from the biased ones can be equally acquired by speakers. This makes sure that other types of structures, once induced due to other constraints, can also be transmitted across generations of leaners.

The above discussion reveals a complicated relation between language and perception.

On the one hand, during cultural transmission of language across multiple generations of learners, individuals' perceptual constraints could favor some structures congruent with the perceptual constraints, thus causing a bias towards those structures. This has been demonstrated in many experimental and simulation studies. For example, some experiments have

shown that the dominant word orders in world's languages are also easier to learn (Culbertson et al., 2012; Fedzechkina et al., 2012; Culbertson and Adger, 2014). Some simulations have also revealed that cultural transmission could amplify small biases for certain structure and make it prevalent in communal languages of later generations (Griffiths and Kalish, 2007; Kirby et al., 2008; Smith, 2011).

On the other hand, other factors, such as different socio-cultural histories could induce distinct language structures. Some modeling studies have demonstrated that socio-cultural interactions could trigger a variety of structural forms, which can be equally acquired and transmitted by generations of language learners (Steels and Belpaeme, 2005; Steels, 2011, 2012). More importantly, as shown in our study, if different structures are more or less functionally equivalent, they can be acquired equally well by speakers, given sufficient rounds of learning. This may diminish the bias for certain structures to a certain extent, and lead to diversity in structural configuration of language.

These two aspects suggest that the actual structures in different languages have arisen as a compromise between both the individual perceptual constraints and the socio-cultural factors (Christiansen and Chater, 2008; Liu, 2014). Such compromise leads to a biased distribution of languages predominantly in certain structures. The mutual influence of individual and socio-cultural factors has been illustrated in some simulation studies of word order bias (Gong et al., 2009) and color naming patterns (Baronchelli et al., 2012).

Our experiment, as an individual learning experiment, could not fully demonstrate the mutual influence of individual and socio-cultural aspects. Nonetheless, it confirmed the influence of individual perceptual constraint (i.e., the perceptual saliency hierarchy) on learning congruent and incongruent orders. It also revealed that given more trials the learning of incongruent orders could reach a similar level to the learning of congruent orders. This suggested that the structures induced by other factors, even though conflicting to the structures favored by perceptual constraints, could still be acquired and transmitted. Both of these findings can shed light on the relation between individual learning and cultural transmission, and contribute to the discussion of the causal factors for the structural diversity in languages (Longobardi and Guardiano, 2009).

Finally, some aspects of this study can be extended in future work. For example, we may recruit pre-language or language-learning children to further diminish the influence of individuals'

prior linguistic knowledge (Saffran et al., 2007; Folia et al., 2010). Compared with visual presentation, auditory presentation better resembles language exchange in everyday life. It may reduce the effect of orthography in language learning, though not eliminating it altogether (Cuskley et al., 2015). In auditory presentation, factors such as memory (Morrison et al., 2014), stress or prosody may also modulate the biases for certain structures.

ETHICS STATEMENT

The experimental protocol was approved by the College Research Ethics Committee of University of Hong Kong. The methods were carried out in accordance with the approved guidelines from the College Research Ethics Committee. Written informed consents were obtained from all participants.

AUTHOR CONTRIBUTIONS

TG and LS designed the research, YL carried out the study. TG and YL analyzed the results. TG, LS, and YL wrote the paper.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01952/full#supplementary-material

REFERENCES

Baronchelli, A., Gong, T., Puglisi, A., and Loreto, V. (2012). Modelling the emergence of universality in color naming patterns. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2403–2407. doi: 10.1073/pnas.0908533107

Braine, M. D. S. (1963). On learning the grammatical order of words. *Psychol. Rev.* 70, 323–348. doi: 10.1037/h0047696

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artif. Life* 8, 25–54. doi: 10.1162/106454602753694756

Brooks, P. J., Kempe, V., and Donachie, A. (2011). Second language learning benefits from similarity in word endings: evidence from Russian. *Lang. Learn.* 61, 1142–1172. doi: 10.1111/j.1467-9922.2011.00665.x

Callaghan, T. C. (1990). "Interference and dominance in texture segregation," in *Visual Search*, ed. D. Brogan (New York, NY: Taylor & Francis), 81–87.

Carter, R., McCarthy, M., Mark, G., and O'Keeffe, A. (2011). *English Grammar Today: An A-Z of Spoken and Written Grammar*. London: Cambridge University Press.

Chater, N., Reali, F., and Christiansen, M. H. (2009). Restrictions on biological adaptation in language evolution. *Proc. Natl. Acad. Sci. U.S.A.* 106, 1015–1020. doi: 10.1073/pnas.0807191106

Christiansen, M. H., and Chater, N. (2008). Language as shaped by the brain. *Behav. Brain Sci.* 31, 489–509. doi: 10.1017/S0140525X08004998

Christiansen, M. H., and Chater, N. (2015). The language faculty that wasn't: a usage-based account of natural language recursion. *Front. Psychol.* 6:1182. doi: 10.3389/fpsyg.2015.01182

Christiansen, M. H., and Ellefson, M. R. (2002). "Linguistic adaptation without linguistic constraints: the role of sequential learning in language evolution," in *The Transition to Language*, ed. A. Wray (Oxford: Oxford University Press), 335–358.

Christiansen, M. H., and Kirby, S. (2003). Language evolution: consensus and controversies. *Trends Cogn. Sci.* 7, 300–307. doi: 10.1016/S1364-6613(03)00136-0

Cinque, G. (1999). *Adverbs and Functional Heads: A Cross-linguistic Perspective*. London: Oxford University Press.

Cinque, G. (2010). *The Syntax of Adjectives: A Comparative Study*. Cambridge, MA: MIT Press.

Cinque, G., and Rizzi, L. (2008). The cartography of syntactic structures. *Stud. Ling.* 2, 42–58. doi: 10.1007/s10936-014-9330-6

Claidière, N., Smith, K., Kirby, S., and Fagot, J. (2014). Cultural evolution of a systematically structured behavior in a non-human primate. *Proc. R. Soc. B* 281: 20141541. doi: 10.1098/rspb.2014.1541

Conway, C. M., and Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends Cogn. Sci.* 5, 539–546. doi: 10.1016/S1364-6613(00)01800-3

Cornish, H. (2010). Investigating how cultural transmission leads to the appearance of design without a designer in human communication systems. *Interact. Stud.* 11, 112–137. doi: 10.1075/is.11.1.02cor

Culbertson, J., and Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5842–5847. doi: 10.1073/pnas.1320525111

Culbertson, J., Smolensky, P., and Legendre, G. (2012). Learning biases predict a word order universal. *Cognition* 122, 306–329. doi: 10.1016/j.cognition.2011.10.017

Cuskley, C., Simner, J., and Kirby, S. (2015). Phonological and orthographic influences in the bouba-kiki effect. *Psychol. Res.* doi: 10.1007/s00426 [Epub ahead of print].

De Boer, D., and Verhoef, T. (2012). Language dynamics in structured form and meaning spaces. *Adv. Complex Syst.* 15:1150021. doi: 10.1142/S0219525911500214

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Lang. Ling. Comp.* 6, 654–672. doi: 10.1002/lnc3.361

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Topics Cogn. Sci.* 19, 603–615. doi: 10.1016/j.tics.2015.07.013

Dominy, N. J., and Lucas, P. W. (2001). Ecological importance of trichromatic vision to primates. *Nature* 410, 363–366. doi: 10.1038/35066567

Dunn, M., Green, S. J., Levinson, S. C., and Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473, 79–82. doi: 10.1038/nature09923

Elman, J. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402_1

Esper, E. A. (1925). A technique for the experimental investigation of associative interference in artificial linguistic material. *Lang. Mono.* 1, 1–47.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., and Wong, P. C. M. (2015). The relationship between artificial and second language learning. *Cogn. Sci.* 40, 822–847. doi: 10.1111/cogs.12257

Evans, N., and Levinson, S. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–492. doi: 10.1017/S0140525X0999094X

Fedzechkina, M., Jaeger, T. F., and Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17897–17902. doi: 10.1073/pnas.1215776109

Ferrer-i-Cancho, R. (2004). Euclidean distance between syntactically linked words. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 70(5 Pt 2):056135. doi: 10.1103/PhysRevE.70.056135

Folia, V., Uddén, J., de Vries, M., Forkstam, C., and Petersson, K. M. (2010). Artificial language learning in adults and children. *Lang. Learn.* 60(Suppl. 2), 188–220. doi: 10.1111/j.1467-9922.2010.00606.x

Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the "critical period" hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 529–534. doi: 10.1073/pnas.012611199

Futrell, R., Mahowald, K., and Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10336–10341. doi: 10.1073/pnas.1502134112

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cogn. Sci.* 29, 737–767. doi: 10.1207/s15516709cog0000_34

Galantucci, B., and Garrod, S. (2010). Experimental semiotics: A new approach for studying the emergence and the evolution of human communication. *Interact. Stud.* 11, 1–13. doi: 10.1075/is.11.1.01gal

Galantucci, B., and Garrod, S. (2011). Experimental semiotics: a review. *Front. Hum. Neurosci.* 5:11. doi: 10.3389/fnhum.2011.00011

Gentner, T. Q., and Goldin-Meadow, S. (eds) (2003). *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press.

Gong, T., Minett, J. W., and Wang, W. S.-Y. (2009). A simulation study on word order bias. *Interact. Stud.* 10, 51–75. doi: 10.1075/is.10.1.04gon

Gong, T., Shuai, L., and Wu, Y. (2013). Multidisciplinary approaches in evolutionary linguistics. *Lang. Sci.* 37, 1–13. doi: 10.1111/brv.12160

Gong, T., Shuai, L., and Zhang, M. (2014). Modelling language evolution: examples and predictions. *Phys. Life Rev.* 11, 280–302. doi: 10.1016/j.plrev.2013.11.009

Griffiths, T. L., and Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cogn. Sci.* 31, 441–480. doi: 10.1080/15326900701326576

Gómez, R. L., and Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends Cogn. Sci.* 4, 178–186. doi: 10.1016/S1364-6613(00)01467-4

Halliday, M. A. K. (1967). Notes on transitivity and theme in english (part II). *J. Ling.* 3, 199–244. doi: 10.1017/S0022226700016613

Haspelmath, M. (2007). Pre-established categories don't exist: consequences for language description and typology. *Ling. Typol.* 11, 119–132.

Healey, C. G. (2000). Building a perceptual visualization architecture. *Behav. Inf. Technol.* 19, 349–366. doi: 10.1080/014492900750000054

Healey, C. G., and Enns, J. T. (1999). Large datasets at a glance: Combining textures and colors in scientific visualization. *IEEE Trans. Vis. Comput. Graph.* 5, 145–167. doi: 10.1109/2945.773807

Healey, C. G., and Enns, J. T. (2012). Attention and visual memory in visualization and computer graphics. *IEEE Trans. Vis. Comput. Graph.* 18, 1170–1188. doi: 10.1109/TVCG.2011.127

Heine, B. (2001). *The Cognitive Foundations of Grammar*. Oxford: Oxford University Press.

Heine, B., and Kuteva, T. (2008). *The Genesis of Grammar: A Reconstruction*. Oxford: Oxford University Press.

Hurford, J. (2007). *The Origins of Meaning*. Oxford: Oxford University Press.

Hurford, J. (2012). *The Origins of Grammar*. Oxford: Oxford University Press.

Kirby, S. (1999). *Function, Selection and Innateness: The Emergence of Language Universals*. Oxford: Oxford University Press.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10681–10686. doi: 10.1073/pnas.0707835105

Kirby, S., Dowman, M., and Griffiths, T. (2006). Innateness and culture in the evolution of language. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5241–5245. doi: 10.1073/pnas.0608222104

Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Liu, H. T. (2010). Dependency distance as a means of word-order typology: a method based on dependency treebanks. *Lingua* 120, 1567–1578. doi: 10.1016/j.lingua.2009.10.001

Liu, H. T. (2014). Language is more a human-driven system than a semiotic system: comment on "Modelling language evolution: examples and predictions" by Tao Gong, Lan Shuai, Menghan Zhang. *Phys. Life Rev.* 11, 309–310. doi: 10.1016/j.plrev.2013.12.008

Longobardi, G., and Guardiano, C. (2009). Evidence for syntax as a signal of historical relatedness. *Lingua* 119, 1679–1706. doi: 10.1016/j.lingua.2008.09.012

Lupyan, G., and Christiansen, M. H. (2002). "Case, word order, and language learnability: Insights from connectionist modeling," in *Proceedings 24th Annual Cognitive Science Society*, eds W. D. Gray and C. Schunn (Mahwah, NJ: Lawrence Erlbaum Associates), 596–601.

Mathesius, V. (1975). *A Functional Analysis of Present Day English on a General Linguistic Basis*, ed. J. Vachek. The Hague: Mouton.

Melin, A. D., Hiramatsu, C., Fedigan, L. M., Schaffner, C. M., Aureli, F., and Kawamura, S. (2012). "Ploymorphism and adaptation of primate color vision,"

in *Evolutionary Biology: Mechanisms and Trends*, ed. P. Pontarotti (London: Springer), 225–241.

Mesoudi, A. (2011). *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago, IL: University of Chicago Press.

Mesoudi, A., and Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philos. Trans. R. Soc. B* 363, 3489–3501. doi: 10.1098/rstb.2008.0129

Misyak, J. B., and Christiansen, M. H. (2012). Statistical learning and language: an individual differences study. *Lang. Learn.* 62, 302–331. doi: 10.1111/j.1467-9922.2010.00626.x

Misyak, J. B., Christiansen, M. H., and Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Front. Psychol.* 1:31. doi: 10.3389/fpsyg.2010.00031

Moeser, D. S., and Bregman, A. S. (1972). The role of reference in the acquisition of a miniature artificial language. *J. Verbal Learn. Verbal Behav.* 11, 759–769. doi: 10.1016/S0022-5371(72)80010-0

Morgan, J. L., and Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *J. Verbal Learn. Verbal Behav.* 20, 67–85. doi: 10.1016/S0022-5371(81)90312-1

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K., Carpenter, H., and Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism* 17, 56–72. doi: 10.1017/S1366728912000715

Morrison, A. B., Conway, A. R. A., and Chien, J. M. (2014). Primacy and recency effects as indices of the focus of attention. *Front. Hum. Neurosci.* 8:6. doi: 10.3389/fnhum.2014.00006

Onnis, L. (2012). "Artificial language learning," in *Routledge Encyclopedia of Second Language Acquisition*, ed. P. Robinson (New York, NY: Routledge), 5–39.

Osborne, J. W., and Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Pract. Assess. Res. Eval.* 9, 1–12.

Pemiss, P., and Vigliocco, G. (2014). The bridge of iconicity: from a world experience to the experience of language. *Philos. Trans. R. Soc. B* 369:20130300. doi: 10.1098/rstb.2013.0300

Petersson, K. M., Folia, V., and Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain Lang.* 120, 83–95. doi: 10.1016/j.bandl.2010.08.003

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychol. Bull.* 133, 227–244. doi: 10.1037/0033-2909.133.2.227

Pullum, G. K., and Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *Ling Rev* 19, 9–50.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reali, F., and Christiansen, M. H. (2009). Sequential learning and the interaction between biological and linguistic adaptation in language evolution. *Interact. Stud.* 10, 5–30. doi: 10.1075/is.10.1.02rea

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. (1993). *Implicit Learning and Tacit Knowledge*. New York, NY: Oxford University Press.

Richerson, P. J., and Christiansen, M. H. (eds) (2013). *Cultural Evolution: Society, Technology, Language, and Religion*. Cambridge, MA: MIT Press.

Robinson, P. (2010). Implicit artificial grammar and incidental natural second language learning: how comparable are they? *Lang. Learn.* 60(Suppl. 2), 245–263. doi: 10.1111/j.1467-9922.2010.00608.x

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

Saffran, J. R., Pollak, S. D., Seibel, R. L., and Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition* 105, 669–680. doi: 10.1016/j.cognition.2006.11.004

Scott-Phillips, T. C., Kirby, S., and Ritchie, G. R. S. (2009). Signaling signalhood and the emergence of communication. *Cognition* 113, 226–233. doi: 10.1016/j.cognition.2009.08.009

Simner, J., Cuskley, C., and Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception* 39, 553–569. doi: 10.1068/p6591

Smith, K. (2011). Learning bias, cultural evolution of language, and the biological evolution of the language faculty. *Hum. Biol.* 83, 261–278. doi: 10.3378/027.083.0207

Smith, K., Brighton, H., and Kirby, S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Adv. Complex Syst.* 6, 537–558. doi: 10.1371/journal.pone.0037744

Steels, L. (2011). Modelling the cultural evolution of language. *Phys. Life Rev.* 8, 339–356. doi: 10.1016/j.plrev.2011.10.014

Steels, L. (2012). *Experiments in Cultural Language Evolution*. Amsterdam: John Benjamins.

Steels, L., and Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: a case study for color. *Behav. Brain Sci.* 28, 469–488. doi: 10.1017/S0140525X05000087

Tabullo, Á, Arismendi, M., Wainselboime, A., Primero, G., Vernis, S., Segura, E., et al. (2012). On the learnability of frequent and infrequent word orders: an artificial language learning study. *Q. J. Exp. Psychol.* 65, 1848–1863. doi: 10.1080/17470218.2012.677848

Tamariz, M., and Kirby, S. (2015). Culture: copying, compression, and conventionality. *Cogn. Sci.* 39, 171–183. doi: 10.1111/cogs.12144

Taylor, J. S. H., Plunkett, K., and Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: an artificial orthography paradigm. *J. Exp. Psychol.* 37, 60–76. doi: 10.1037/a0020126

Tily, H., Frank, M. C., and Jaeger, T. G. (2011). The learnability of constructed languages reflects typological patterns. *Cogn. Sci. Soc.* 33, 1364–1369.

Tomlin, R. S. (1986). *Basic Word Order: Functional Principles*. London: Croom Helm.

Treisman, A. (1985). Preattentive processing in vision. *Comput. Vis. Graph Image Process.* 31, 156–177. doi: 10.1016/S0734-189X(85)80004-9

Yip, P.-C., and Rimmington, D. (2004). *Chinese: A Comprehensive Grammar*. London: Routledge.

Zhu, Q. (2005). *Modern Chinese Practical Grammar Analysis*. Beijing: Tsinghua University Press.

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.