Cultural Transmission and Evolution of Melodic Structures in Multi-generational Signaling Games

Massimo Lumaca*·**
SISSA International School for
 Advanced Studies
Aarhus University

Giosuè Baggio†
SISSA International School for
 Advanced Studies
Norwegian University of Science
 and Technology

Abstract   It has been proposed that languages evolve by adapting to the perceptual and cognitive constraints of the human brain, developing, in the course of cultural transmission, structural regularities that maximize or optimize learnability and ease of processing. To what extent would perceptual and cognitive constraints similarly affect the evolution of musical systems? We conducted an experiment on the cultural evolution of artificial melodic systems, using multi-generational signaling games as a laboratory model of cultural transmission. Signaling systems, using five-tone sequences as signals, and basic and compound emotions as meanings, were transmitted from senders to receivers along diffusion chains in which the receiver in each game became the sender in the next game. During transmission, structural regularities accumulated in the signaling systems, following principles of proximity, symmetry, and good continuation. Although the compositionality of signaling systems did not increase significantly across generations, we did observe a significant increase in similarity among signals from the same set. We suggest that our experiment tapped into the cognitive and perceptual constraints operative in the cultural evolution of musical systems, which may differ from the mechanisms at play in language evolution and change.

1 Introduction

Over the past twenty years, the evolution of language and other symbolic systems has been understood from a broadly Darwinian perspective, in which languages are viewed as complex adaptive systems [9, 16]. During cultural transmission, languages change as the result of pressures to adapt to the constraints imposed, among other factors, by the human brain. Computational and experimental

research has indeed suggested that, as a result of transmission, languages become more regular (on several dimensions) and easier to learn and transmit [35, 37]. One may ask whether and to what extent similar principles extend to the evolution of melodic systems, such as music and vocalization [53].

Universal acoustic and auditory mechanisms may explain the convergence on a narrow set of melodic features found across cultures. Although the idea is not new [74, 78], only a few studies have addressed it by means of laboratory experimentation [79].

1.1 Iterated Learning and Beyond

In a study by Verhoef [79], the emergence of combinatoriality was investigated using an artificial sound system transmitted in the laboratory. A set of whistled sounds was presented to the first participant in a diffusion chain. The participant had to learn and reproduce the set as accurately as possible. The output was then presented through a computer to the second participant, who underwent the same procedure, and so on for several iterations (*generations*). This paradigm is a variant of iterated learning [35]. In the course of transmission, structural regularities emerged. In the last few generations, the set of signals was structurally rich, compressed, and easier to learn and reproduce.

Despite the effectiveness of iterated learning in modeling the emergence of structural regularities in artificial miniature languages (see [36] for an overview), the model in use by Verhoef [79] requires the direct intervention of the experimenter in order to filter out redundant signals from training sets: monomelodic signal sets were otherwise expected to emerge. These interventions were intended to mimic the adaptive pressure imposed by communication [63], which is by design absent in this version of iterated learning (but see [38] for a new procedure; see [52] for a critical discussion).

Here, we investigate the evolution in the laboratory of a miniature artificial tone system of meaningful sequences built out of discrete variable tones. We assess whether cognitive and perceptual constraints shape auditory codes so that they gradually acquire features that may enhance acquisition and memory retention [50]. To this end, we test whether artificial tone systems would evolve toward more compressible and simpler forms [71], using measures of *compositionality* [35] and *melodic compression* (see [13] for a similar measure with visual stimuli). Next, we test whether tone systems would evolve so as to embody perceptual principles of organization, such as *auditory grouping* [18].

To address these issues we introduce three major innovations as compared to previous iterated learning studies. First, we use a system of signals based on discrete pitches (a *scale*) [70]. This is cognitively relevant for both music [53] and speech intonation perception [48]; in addition it allows us to introduce quantitative (as opposed to qualitative [79]) analyses to track the evolution of melodic structures. Second we present and apply a model of cultural evolution (*signaling games*) in which the transmission and acquisition of codes occurs via repeated rounds of interaction between senders and receivers. While this is not a novel approach in language evolution research [9, 75] (see especially Nowak and Baggio [52]), only one study has exploited it in the auditory domain (Lumaca and Baggio [46]). We expect structural regularities to arise spontaneously and gradually as a result of coordination and communication, possibly reflecting processing constraints. Third, we use (affective) meanings associated to each melodic signal, here represented by facial expressions. The use of artificial tone systems endowed with a semantics makes our research relevant for investigating the evolution of symbolic vocal or melodic systems [53]. It is indeed widely accepted that speech intonation can convey a broad variety of meanings, from affective to pragmatic, while it still remains unclear whether similar arguments might be applied to musical melody [25]. There is broad agreement among musicologists and cognitive scientists that musical phrases (and smaller or larger units of musical discourse) can be used, and have historically been used, to convey a variety of types of meaning, including affective meaning [15, 40–42]. In the context of this experiment, we define a *signal* as a (possibly melodic) sequence of discrete tones, and a *code* as a set of mappings between signals and meanings in the minds of the two players [51]. We chose emotions conveyed by facial expressions as meanings, given the links between facial expressions and music-evoked emotions found in behavioral and neural research (see [30, 54] and especially [43]). This will allow us, following

Kirby et al. [35], to explore whether systematicity in the mapping of signals to meanings, or compositionality, arises in the organization of symbolic pitch patterns.

1.2 The Present Study

In signaling games [45, 66], a sender and a receiver coordinate to share information on states of affairs. In each signaling trial, the sender has private access to a state, and uses a signal to inform the receiver on the identity of the state. The receiver must in turn take action. If the action does not match the observed state, communication fails. Crucially, in signaling games there is no mapping of signals to states given to or negotiated by the participants prior to the start of a game. The sender and receiver must coordinate, through trial and error or other learning mechanisms, to develop a common code. The extent to which each player adapts his own signal-state mappings to those of the other player reveals the division of coordination labor between the two players.

Bidirectional negotiation of cultural material between peers or generations is one aspect of cultural transmission [62], which may also hold in the evolution of auditory symbolic systems. In the standard versions of iterated learning [35, 79], there is no coordination process occurring between players, and asymmetry of transmission is fixed: no learner can affect the behavior of the previous generation. This is not the case in our model, where receivers interact with senders and can dynamically negotiate a signaling system. In typical instances of cultural transmission, a net flow of information from senders to receivers is necessary for the maintenance of the symbolic system over time. Moreno and Baggio [51] show that this condition is achieved if sender and receiver play with fixed roles throughout a game. In this condition, most coordination labor falls to the receiver. It is therefore the sender's code that tends to become the common code. Signaling games with fixed roles are a viable model of the transmission of symbolic systems in vertical diffusion chains [52].

In our experiment, the states (possible messages) to be denoted by signals are emotions, shown as facial expressions, and signals are sequences of five pure tones drawn from the Bohlen-Pierce (BP) scale [47]. The group of signals belonging to the same code, and denoting different emotions, will be referred to from now on as a *signal set* (or *melodic set*). The BP scale is used to prevent players from exploiting entrenched musical memories in mapping properties of certain tone sets (e.g., the major mode in the Western diatonic scale) with specific emotions (e.g., happiness). The participants'

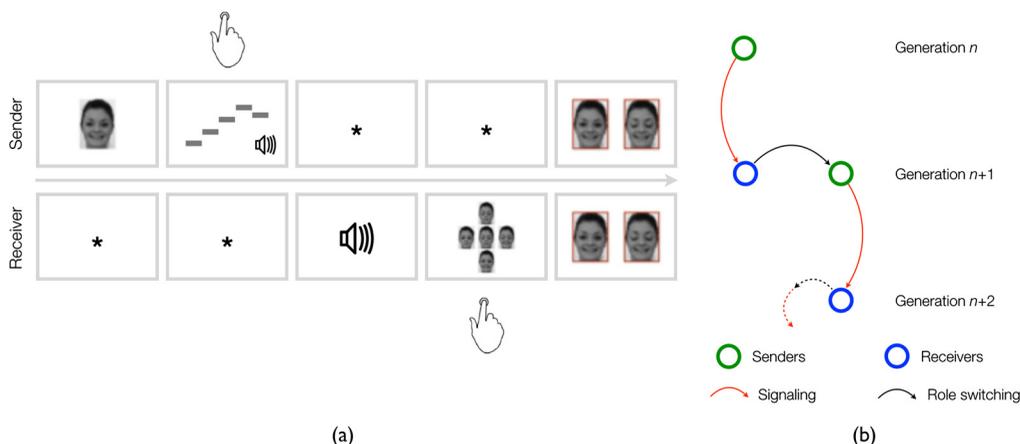(a)   (b)

Figure 1. (a) A trial of a signaling game with five emotions (three simple, two compound) as states (or messages) and five-tone sequences as signals. Emotions are shown as expressive faces. Tones are drawn from the Bohlen-Pierce scale. The top and the bottom rows show what the sender and receiver view on their screen, respectively. Time flows from left to right. Feedback is positive if the facial expression viewed by the sender and the one chosen by the receiver match, and negative otherwise. With positive feedback, facial expressions are shown to both participants simultaneously in a green frame, while a red frame is used for negative feedback. (b) Structure of multi-generational signaling games. The receiver in one game (generation *n*) is the sender in the next game (generation *n* + 1).

task is to develop and finally agree on a mapping of signals to states over successive signaling rounds (Figure 1a). The initial code (or *seed*) is produced by the experimenters. Seeding is a common technique in experimental studies of cultural transmission [26, 81]. The seed is used to train the first participant (sender) of each diffusion chain (generation 1). A signaling game is then played between the first and second participants (generation 2). These two participants play respectively as the sender and receiver, with their roles held constant throughout a session. At the end of each game, the receiver switches his role and becomes the sender in the next game, now playing with a naive individual as a receiver (generation 3) (Figure 1b). Multi-generational signaling games are thus played sequentially and are organized as in a vertical diffusion chain consisting of eight generations.

As a first result, we expect to further validate signaling games as a model of the cultural transmission of symbolic systems [46, 51, 52]. We assessed the direction of information flow in the chains using two indexes: *role asymmetry* and *coordination* [51]. When players share most mappings at the end of a game, so that coordination in practice occurs, and the receiver has adjusted his mappings more frequently to those of the sender (there is *asymmetry* in the coordination process), then we can conclude that information has flowed forward in the chain, from senders to receivers.

In addition, we test how constraints on compression of information [10] and on melodic perception and memory [17] shape the structure of tone sequences during transmission. First, we apply measures of compositionality and melodic compression. These measures can assess how similar are signals of the same set (melodic compression), and whether systematic relations exist between signal segments and meanings (compositionality). These two constraints are thought to drive signaling systems toward compressed, easier-to-learn forms [35, 71]. Enhanced processing may be produced by peculiar structural features embodied in melodic segments, or grouping features. To this purpose, changes in the melodic structure will be assessed by testing three gestalt principles of sequence organization: proximity, good continuation, and equivalence. All measures are formally defined below.

Finally, to test changes in the learnability of the code we used an index of transmission (T) and an index of innovation (I). Communication pressure introduced by signaling games is expected to counteract learning demands, maintaining a rich set of diversified signals.

2 Material and Methods

2.1 Participants

Sixty-four participants took part in the study (29 female, mean age 24.3, range 19–32). All had normal hearing and normal or corrected-to-normal visual acuity, and no formal musical training. Signaling games were played in an experimental room on two computer terminals facing each other. Screens were aligned back to back, making it impossible for either player to see their partner and their screen. Each participant was provided with a full-size standard PC keyboard and stereo headphones. Participants were not allowed to communicate verbally or otherwise besides the signaling game itself. The experimenter was always present in the room.

2.2 Diffusion Chains

Signaling games were organized in eight vertical diffusion chains of eight generations each. At the end of each game, the receiver in generation n became the sender in the next game, playing with a new receiver in generation $n + 1$. The sender was instructed to transmit the musical code as he recalled it from the previous game.

2.3 States

The states or possible messages, denoted by signals in signaling games, were five *emotion categories* of different complexity: three *basic* and two *compound emotions*, represented as facial expressions (Figure 2). Although the perception and production of compound facial expressions is consistent with the subordinate basic categories they are built upon [22], they are perceived as independent expressions. Following Ekman and Friesen [23], facial expressions of basic emotions (peace, joy,
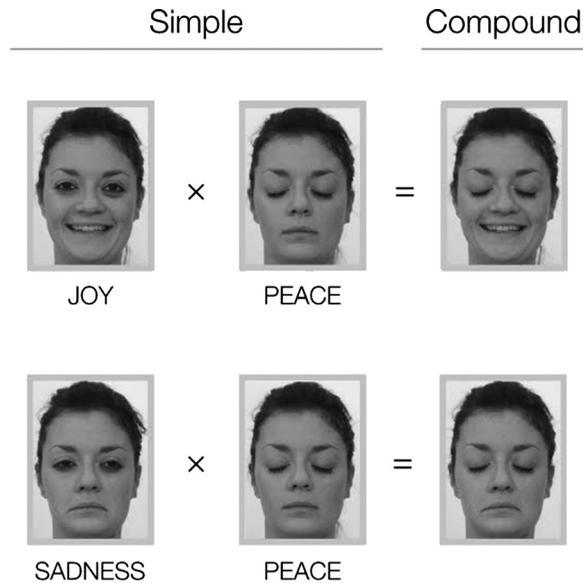
Figure 2. Stimulus set of the study on emotion expression. States are faces expressing either simple (sadness, joy, peace) or compound emotions (peace × joy, peace × sadness). Stimuli for compound emotions are constructed by merging the upper portion of peace with the lower portion of joy or sadness.

and sadness) were analyzed in two facial regions (or meaning dimensions): *upper face* and *lower face*. Compound emotions were built using the upper face features of peace and the lower face features of joy (peace + joy) and sadness (peace + sadness). For the purposes of compositionality analyses, the upper face was coded using two variants (*open* or *closed* eyes), and the lower face in three variants (mouth corners *up*, *straight*, and *down*).

2.4 Signals
The five constituent tones of the signals were drawn from the Bohlen-Pierce (BP) scale [47]. In the equal-tempered version, a tritave (3 : 1 frequency ratio) is logarithmically divided into 13 equal steps (see Figure 1 in the Online Supplement at http://www.mitpressjournals.org/doi/suppl/ 10.1162/ARTL_a_00238), larger in size than the corresponding Western semitones (146 cents versus 100 cents). This makes BPS a *macrotonal* scale. Here "macrotone" is the term used to define the smallest units of pitch (i.e., of size larger than a semitone). Pitches of the BP scale are defined by the following equation: $F = k \times 3^{n/13}$, where k is the reference pitch frequency, and n is the number of steps on the scale. We set $k = 220$ Hz with n equal to 0, 4, 6, 7, and 10 so as to maximize the number of low integer frequency ratios between any combination of tones. Sounds were sine waves 500 ms long, with 50-ms fade-in and -out and 50 ms inter-tone interval (see Figure 2 and related audio material in the Online Supplement). Adjacent numeric keys (1 to 5) of the computer keyboard were used to produce the five sounds. Stimuli were delivered through headphones at 80 dB. No signal space was set a priori, in view of the analysis of compositionality (see below).

2.5 Procedure
Prior to the start of a game, each player in a pair was independently trained to associate each facial expression with one simple or compound emotion, in three blocks. In the first and second blocks, respectively, facial expressions for simple and compound emotions were randomly presented at the center of the screen to the participant, one at a time. Given each facial expression, the participant had to choose the appropriate emotion (e.g., peace + joy) among the options on the screen, using numerical keys of the computer keyboard (1 to 5). Feedback followed, in which the correct response for the facial expression was shown, highlighted in green. During a test in block 3, the entire set of

facial expressions was presented in a row, but this time without feedback. Four correct responses for each facial expression were necessary to complete the block.

At the end of training, participants moved to a separate testing room to play the signaling game. At that stage, each player received written instructions about the main experimental task. In the first part, they were told that the goal of the task was for players to develop a communication system made of sounds and emotional expressions through repeated interactive trials, and that no other form of communication, verbal or otherwise, was allowed besides the signaling games. Specific instructions for the sender and receiver followed, in which the structure of the game was described in detail, step by step. The receiver was informed of his role in the next game as sender. Figure 1a shows the structure of a single signaling trial. The sender was presented with a randomly selected facial expression from one actor's photo set (e.g., joy; 5-s duration); a blank screen followed. Next, the sender was asked to generate a five-tone sequence from the pool of five tones available, using the keyboard, to signal to the receiver the state he had seen. All tone sequences were isochronous: melodic patterns (pitch), but not rhythm (timing), were controlled by senders. Across trials, the order of the tones was randomized in the auditory choice sequence presented to senders, to prevent them from creating associations between states and positions in a fixed sequence. Thus, senders had to discover the key-tone mapping anew for each trial to produce the intended tone sequence. Unheard by the receiver, the sender could try any combination of five tones at will; the mapping of tones to keys was held constant for each within-trial attempt (each signal actually sent). The signal was then transmitted to the receiver, who listened to it via headphones. In turn, the receiver was asked to choose one of the five facial expressions shown on the screen (i.e., the one he thought the sender had seen) by using the keys labeled 1–5 on the keyboard. The order of the facial expressions as shown to the receiver was randomized over trials. A simultaneous feedback (3 s) to both players followed, displaying the expression the sender had seen (in a green frame), and the one the receiver had chosen (in a green frame if correct; in a red frame if incorrect). The end of the game was set to 50 correct trials, with one break at 25 correct trials.

2.6   Seeding

We designed four sets of five-tone sequences (Online Supplement Figure 3) characterized by controlled levels of compositionality ($M = 0.01$, SEM $= 0.002$, range [0:1]), contour smoothness (Shannon entropy, $M = 1.19$, SEM $= 0.04$, range 0:2.32), and proximity (absolute mean interval size, $M = 12.52$, SEM $= 0.62$, range 0:28). Prior to the beginning of a session, the first sender was trained with the seeding stimuli in five blocks of increasing complexity and duration. In each block, he was instructed to produce a five-tone sequence for a given state and the correct signal for states presented in previous blocks. Five correct sequences per state were necessary to move on to a new block or complete the training phase.

2.7   Data Analysis

The aim of this study is to identify and quantify the evolution of melodic structures and regularities during transmission. Nonparametric Wilcoxon signed-rank tests [65] were used to compare data points between generations. To examine cumulative code changes over time, we analyzed the data by means of linear mixed-effects regression models [82] in R Studio (R Studio Team, 2015) with lme4 [4]. The same approach was used in previous research on iterated learning [79]. The dependent variables were structural properties of the musical signals: melodic proximity (mean interval size and interval distribution), symmetry (melodic transformations, such as retrograde and inversions), and contour smoothness (mean contour entropy); moreover, we measured changes in information compression: here, compositionality (systematicity in mapping signals to meanings) and melodic compression (melodic similarity within a signal set). These measures are formally defined below. The dependent variables (y) were modeled as a function of generation (fixed effect), with random intercepts (by-chain variation in y) and random slopes for generation (by-chain variation in the slope of generation). For each dependent variable, we tested a full model against a null model excluding the

effect of generation, using a likelihood ratio test. Indexes of model efficiency (asymmetry, etc.) were analyzed using one-sample Wilcoxon tests of the null hypothesis that they are not significantly different from 0. To include a baseline measure of change, we carried out a separate set of analyses in which the original data were shuffled ($n = 1000$ times). Dotted lines in the figures show these baselines, that is, the values toward which the evolution of tone sequences would tend if it was driven by chance. To test the direction of change (random versus driven), for each measure we ran one-sample Wilcoxon tests between the codes produced in the last generation (G7) and the relative median baseline value. The significance level for all analyses was $\alpha = 0.05$. Pairwise tests between generations were Bonferroni-corrected at $\alpha = 0.05/5 = 0.01$.

2.8   Melodic Similarity

The similarity between pairs of musical signals produced by adjacent generations, and denoting the same emotion, was measured using a modified version of the edit distance [29], here the number of string elements shared between two strings of equal length. In our analyses, the strings were either five-tone sequences (*tone distance*) or their contour transforms, that is, the *constants*, *ups*, and *downs* of melodic intervals, independent of their size (*contour distance*), normalized in the range $0 : 1$. Previous studies have shown that melodic contours are retrieved more accurately than the exact pitch sequence [20].

2.9   Asymmetry and Coordination

In the first step of data analysis, signal-state mappings in each trial were determined separately for the two players. For each state, we identified a *coordination point*, the trial from which sender and receiver use the same mapping consistently until the end of the game (barring random errors) [51]. More specifically, we first identified a block of five consecutive trials (as many as the semantic states) where participants were consistently using the same mapping. We then searched backward in the trial sequence for the first trial in which coordination de facto occurred, that is, where players use the same code, possibly despite their being unaware of that. This point divides the trial sequence for a given state into two portions: the first portion, where players attempt to coordinate (*coordination phase*), and the second portion, where the code is shared (*communication phase*). Hence, two indexes of model efficiency were used: first, *asymmetry*, or the difference in the number of code changes introduced by the sender (S) and by the receiver (R) during coordination, divided by the total of code changes: $A = (S - R)/(S + R)$. Asymmetry ranges from -1 (the receiver adapts his mappings to the sender) to 1 (vice versa), with a single value calculated for each pair (i.e., each game). Second, *coordination*, measured for a pair as the mean similarity between signals of corresponding emotions used by players at the end of a game. Two values for each pair were derived using the actual tone and contour distances. The values range from 0 (no coordination) to 1 (shared code).

2.10   Transmission and Innovation

The *tone* and *contour* similarity measures (modified edit distance) between sequences denoting the same emotion, produced by players of adjacent generations, are indexes of the faithfulness of transmission (a between-player measure). A code that is faithfully transmitted has values close to 1. Innovation is measured as $1 - D$, where D is the distance (tone and contour) between the tone sequence for one emotion learned by a receiver in one game, and the sequence produced in response to the same emotion by the same player (now the sender) in the next game (within-player measure). High values of innovation are associated with restructuring of the musical code by the same player. Note that, in games where coordination is partial, the melodies received and the ones learned by the receiver do not fully match. In that case, though related, the two indexes do not mirror each other. Values close to 1 suggest that the codes were restructured between games. Four values (two for transmission, two for innovation) were measured for each melodic set in the experiment.

2.11 Proximity

A reliable measure of pitch proximity is the absolute mean interval size, here computed for each set of tone sequences used by players. First, we computed the absolute mean interval size for each tone sequence. Then, we averaged the output value across tone sequences of the same set. Low values are associated to sets with more proximal melodies, and high values to those with more distant melodies. The actual statistical distribution of unison, small (1 or 2 macrotone(s) in size) and large intervals (3, 4, 6, 7, 10 macrotones in size) was also calculated for each melody, and averaged across signals of the same set. This metric provides a further, distinct measure of melodic size.

2.12 Measure of Mirror Forms

We tested the evolution of signals towards three types of transformations: *retrogrades*, *inversions (mirror forms)*, and *retrograde inversions* [19]. In retrogrades, the order of tones of the original sequence (e.g., ABCDE) is reversed in the new sequence (EDCBA). In inversions, the direction of the intervals is reversed, so that an ascending interval in a sequence (e.g., CDCDE) becomes descending in the transformed sequence, and vice versa for any descending intervals (CBCBA). In retrograde inversions, an inversion is carried out first and is then reversed as in retrograde transforms. The presence of each transformation in the data was calculated using the mean value in a similarity matrix excluding the main diagonal. The matrix contains pairwise similarity values between elements of the set and the relative transformation.

2.13 Measure of Contour Smoothness

The Shannon entropy [64] of contour transforms was obtained with the following equation:

$$H(X) = -\sum P(x_i) \log_2 P(x_i)$$

where X is the string of constants, ups, and downs and $P(x_i)$ is the probability that element i occurs in that place (estimated from the frequency of the element and the alphabet in use). A single entropic value was computed for each tone sequence, which represents the complexity of the signal surface.

2.14 Compositionality

Compositionality is a measure of how predictably signals and meanings are associated. The compositionality of each melodic set was computed using the information-theoretic tool RegMap (for details see [72]). Following Cornish et al. [14], signals from the last pair of each chain were partitioned into segments, such that elements could be reliably associated to meaning dimensions. As a result of this, signals were segmented into units of two and three tones. We applied this segmentation to the entire pool of signals produced in the experiment. Then, we followed Tamariz [72] and computed the conditional entropy of any possible combination of signal elements and meanings in 1000 randomizations, obtaining a partial RegMap. A partial RegMap specifies to what degree a signal element can predict a meaning element. Last, a single RegMap value for the entire code was computed (range 0 : 1), summarizing the compositionality of the set of signals in a pair.

2.15 Melodic Compression

The melodic compression measures the similarity between signals of the same set [13]. It was computed by taking the mean similarity between signals of the same melodic set (and relative contour transforms). It ranges from 0 (different signals or contour profiles) to 1 (monomelodic or monocontour signal sets).

3   Results

Our results, presented in greater detail below, suggest that a system of tone sequences is restructured when it is transmitted in diffusion chains. Proximal tone material emerged in each set of sequences, featuring contours bound by symmetry relations. A progressive compression in the repertoire of signals was revealed by a gradual increase in similarity among signals of the same melodic set. The evolution of signaling systems from an unstructured state was characterized by early significant changes, followed by incremental variation in the same direction in downstream generations.

3.1   Asymmetry and Coordination

Asymmetry was negative and significantly different from 0 (median $= -0.73$, $n = 56$, $Z = 6.140$, $p < 0.001$). There is a division of labor between players throughout a game, with a tendency for the sender to maintain his initial code and for the receiver to adjust his mapping during coordination. Coordination as measured using tone distance was significantly different from 0 (median $= 1$, $n = 280$, $Z = 16.47$, $p < 0.001$), and so was coordination as measured by contour distance (median $= 1$, $n = 280$, $Z = 16.50$, $p < 0.001$). This confirms previous results that an agreement on shared semantic conventions can be achieved in this version of signaling games [46, 51, 52].

3.2   Transmission and Innovation

Figure 3 shows the evolution of transmission and innovation as measured by tone (Figure 3a) and contour distances (Figure 3b). Using the latter measure, we found an increase in transmission between the second and third generation ($n = 40$, $Z = -3.653$, $p < 0.001$), with a concurrent decrease in innovation ($n = 40$, $Z = -3.665$, $p < 0.001$). However, no significant changes occurred between the third and the last generation (transmission, $n = 40$, $Z = -0.78$, $p = 0.44$; innovation, $n = 40$, $Z = -0.78$, $p = 0.44$). Similar results were obtained with tone distances. Despite high values, from the third game onward transmission was rarely ever perfect and equal to 1 (mean contour distance $= 0.80$, SD $= \pm 0.02$), while innovations introduced by participants remained frequent also in successive generations (mean contour distance $= 0.19$, SD $= \pm 0.02$).
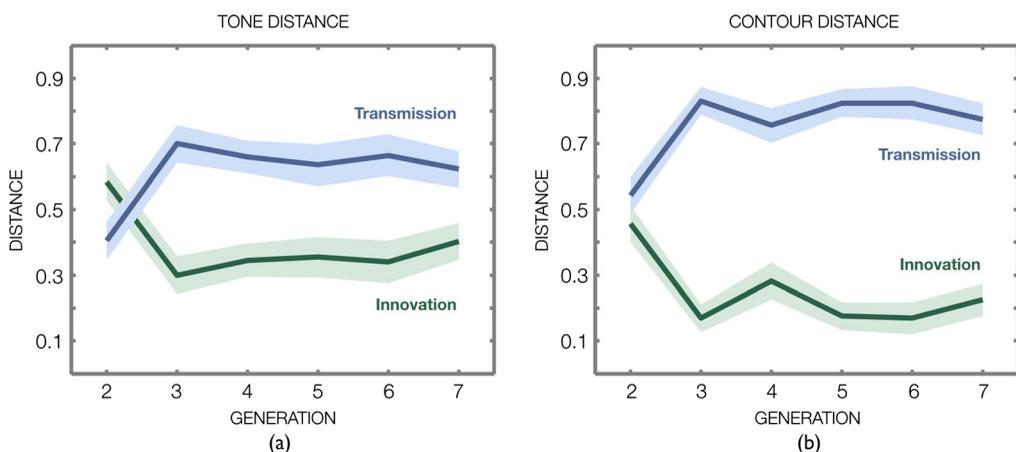

Figure 3. Changes in transmission and innovation over generations as measured by (a) tone and (b) contour distance. The dark line shows the mean value for each generation across chains; the shaded area represents SEMs. *Transmission* is the modified edit distance between codes produced by senders of adjacent generations (*between-players* measure). *Innovation* is calculated as $1 - D$, where D is the distance between the set of codes learned and the set of codes produced in the following game by each participant (*within-player* measure).
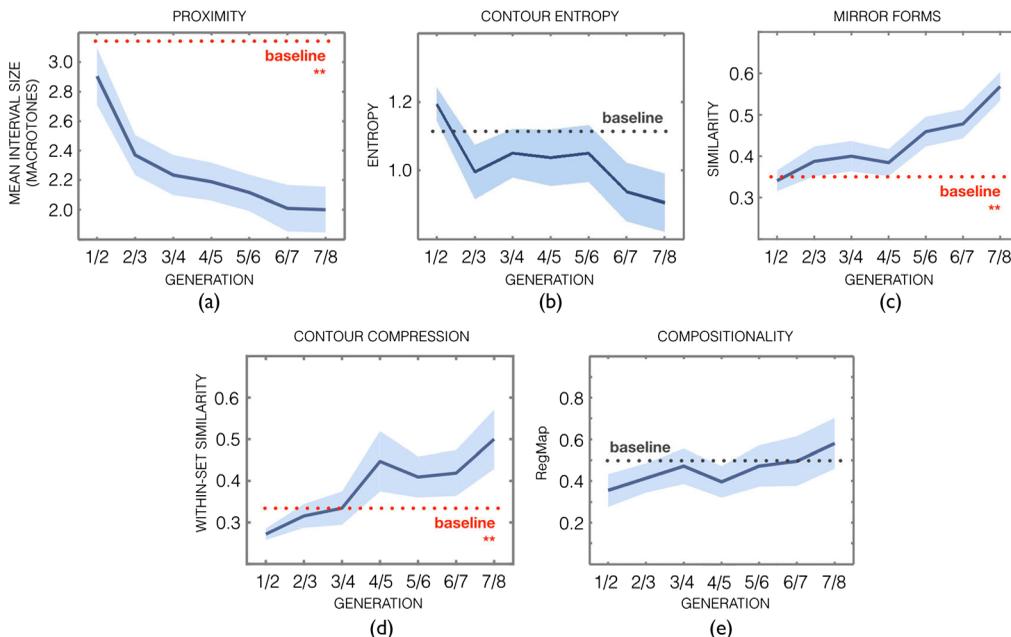
Figure 4. Temporal evolution of structural regularities (a–e) across generations. The dark line is the average value computed for each generation across chains; the blue-shaded area represents the SEM.

The observed increase in transmission fidelity between the second and the third generation suggests that some critical changes in code structure occurred in the output of generation 2. From this point on, moderate changes occurred, as indicated by lower innovation values. The specific direction of such changes is described below (see Online Supplement Figure 4 for an example of evolving codes).

3.3 Proximity

An initial, nonsignificant decrease in interval size ($n = 40$, $Z = -1.838$, $p = 0.06$) was followed by gradual yet robust change in the same direction in successive generations (g2/g7, $n = 40$, $Z = -2.456$, $p = 0.01$) (Figure 4a). This suggests an evolution of signals towards proximal melodies. The cumulative increase of small interval sequences is supported by linear mixed model analysis. The likelihood ratio test of the full model with proximity as test variable against a null model was significant ($\chi^2(1) = 8.15$, $p = 0.004$).

3.4 Interval Distribution

To test whether changes in proximity are due to a progressive increase in the frequency of mono-tone melodies, we performed a two-tone transition analysis. Comparisons of the first and last generation, when generation 1 was included (g1/g7, $n = 80$, $z = -0.084$, $p = 0.93$, Wilcoxon) or excluded from the analysis (g2/g7, $n = 80$, $z = -1.454$, $p = 0.15$), did not show changes in the proportions of unison (0-macrotone intervals). In support of this, the frequency of horizontal melodic shapes represents only a fraction of the signals produced by players. Patterns with changing smoothing features were best represented (Online Supplement Figure 6). These findings suggest that the emergence of proximal intervals could explain the compression in size of signals. Consistent with this idea, we also observed a significant effect of generation on the percentage of large ($\chi^2(1) = 9.60$, $p = 0.001$) and small intervals ($\chi^2(1) = 4.39$, $p = 0.03$), which respectively decreased and increased over time. No effects of generation were found for unisons ($\chi^2(1) = 1.21$, $p = 0.26$).

3.5 Melodic Transformations

A systematic rearrangement of regular patterns occurred through operations performed by players on signals to produce symmetrically related forms. A significant emergence over time of mirror forms or inversions was observed, whether we included the first generation (g1/g7, $n = 80$, $Z = -4.515$, $p < 0.001$) or excluded it from the analyses (g2/g7, $n = 80$, $Z = -3.983$, $p < 0.001$) (Figure 4c). A different pattern was observed for retrogrades (g1/g7, $n = 80$, $Z = -1.835$, $p = 0.06$; g2/g7, $n = 80$, $Z = -1.188$, $p = 0.23$) and retrograde inversions (g1/g7, $n = 80$, $Z = -1.032$, $p = 0.30$; g2/g7, $n = 80$, $Z = -1.860$, $p = 0.06$) (Online Supplement Figure 5). The cumulative increase of mirror forms was confirmed by comparing a linear mixed effect model with inversion as test variable against a null model excluding the fixed effect ($\chi^2(1) = 7.27$, $p = 0.006$). This highlights the importance of symmetrical structures in the evolution of melodic material.

3.6 Contour Smoothness

Smoother melodic lines were also expected to emerge over generations. The Shannon entropy of the melodic contours of tone sequences was used to test this hypothesis. More fragmented contours have higher entropy values than smoother signal surfaces. Figure 4b shows changes in this measure across generations. The decrease between generations 1 and 2 was not significant ($n = 40$, $Z = -1.723$, $p = 0.08$), and the same applies to generations 2 to 7 (g2/g7, $n = 40$, $Z = -1.191$, $p = 0.23$). A decrease of entropy was found using a Wilcoxon signed-rank test (comparing g1 and g7 data, $n = 40$, $Z = -2.455$, $p = 0.01$). However, likelihood ratio tests of the full model with contour entropy as test variable against a null model excluding the fixed effect showed that generation does not affect the change here ($\chi^2(1) = 2.87$, $p = 0.09$). This indicates that the initial melodic material, which mostly consisted of elaborate contours, was replaced by more regular patterns (Online Supplement Figure 6), possibly with different factors at work in producing the observed result (see Section 3.9 below).

3.7 Melodic Compression

Figure 4d shows a gradual increase in similarity across signals of the same melodic set in terms of relative contours. A significant difference between the first and the last generation was found using a Wilcoxon signed-rank test, both including the first generation (g1/g7, $n = 8$, $z = -2.380$, $p = 0.01$) or excluding it in the comparison (g2/g7, $n = 8$, $z = -2.028$, $p = 0.04$). The linear mixed effect analysis confirms a progressive change across generations ($\chi^2(1) = 9.11$, $p = 0.002$). Similar results were obtained when compression was measured on tone sequences ($\chi^2(1) = 6.14$, $p = 0.01$), with a gradual increase between the first and last generation (g1/g7, $n = 8$, $z = -2.533$, $p = 0.01$; g2/g7, $n = 8$, $z = -1.065$, $p = 0.28$).

3.8 Compositionality

Previous results on compression indicate that variability in melodic sets decreases as melodic signals of the same set become similar to one another. This might be explained by the reuse of fewer smaller subsequences to produce whole sequences. Possibly, the combinatorial recycling of these subunits is not random but tailored to the structure of the meanings that should be conveyed. This feature is known as *compositionality*, and its evolution is shown in Figure 4e. The increase in compositionality was not significant, as assessed by a Wilcoxon signed-rank test (g1/g7, $n = 16$, $z = -1.680$, $p = 0.093$) and linear mixed effect analysis ($\chi^2(1) = 3.20$, $p = 0.07$).

3.9 Introducing Baselines

To test whether the observed structural regularities evolved by random drift to mean or baseline levels, rather than by learning and processing constraints in participants, we ran one-sample Wilcoxon tests between original and reshuffled data in each measure ($n = 1000$). While all structural features started from a randomlike state ($p > 0.05$ for all comparisons between original generation 1 data

and shuffled data), the proximity ($n = 8$, $p = 0.02$; baseline $B = 3.46$), mirror forms ($n = 8$, $p <$ 0.01; $B = 0.35$), and melodic compression (tone sequence: $n = 8$, $p = 0.02$; $B = 0.2$; contours: $n = 8$, $p = 0.02$; $B = 0.33$) were found to be significantly different from baseline at the last generation (Figure 4a, c, and d). No significant difference was instead found at the last generation in retrogrades ($n = 8$, $p = 0.12$; $B = 0.35$), retrograde inversions ($n = 8$, $p = 0.43$; $B = 0.35$), contour entropy ($n = 8$, $p = 0.16$; $B = 1.12$), and compositionality ($n = 8$, $p = 0.57$; $B = 0.5$) relative to baseline values. This more stringent analysis against the baseline shows that code evolution in a specific direction, as opposed to random drift, was found for two melodic properties of signals (proximity and symmetry), and for one key requisite of human learning (information compression). This, of course, does not rule out the possibility that random states may have been functional for participants.

4   Discussion

Our results suggest that an artificial system of tone sequences endowed with semantics tends to be regularized in nontrivial ways when it is transmitted across generations. We observed the emergence of structural features that may promote learning and memory retrieval. In particular, proximal melodies with symmetric tone pattern appeared over generations. Moreover, the similarity between melodic segments increased from the first to the last generation, whereas changes in compositionality were negligible and remained around baseline levels.

4.1   Signaling Games as a Model of Cultural Transmission

The initial sharp increase in transmission observed in our data set contrasts with the gradual changes found in previous iterated learning studies [35, 79]. The bidirectional negotiation of the code that takes place during transmission in signaling games may well account for this. Social interaction and repeated communication with the partner may boost the receiver's learning [73], leading to a rapid convergence to a shared symbol-state mapping [69]. As a result, a more faithful reproduction of the code is expected after a few generations.

The code was rarely perfectly reproduced in subsequent generations. Innovations, either deliberate or introduced by memory erosion, maintained tone sequences in a persistent dynamic state and accumulated gradually over generations. High-fidelity transmission and low levels of innovation are two properties of cumulative cultural evolution, which is thought to have a role in the origins of linguistic and musical behavior [8]. In agreement with earlier work [51, 52, 46], receivers changed their mapping of states to signals more often than senders to achieve coordination in a game. This resulted in a net (vertical) flow of information from senders to receivers, and from the first to the last generation of the diffusion chains.

Cumulative transmission and asymmetry in information flow, as found in our study, are essential traits of cultural evolution [76]. These properties, together with the absence of artificial manipulations or filtering of evolving codes [35, 79], make signaling games a viable laboratory model of the cultural evolution of auditory tone systems, and a possible complementary approach to classical iterated learning [52].

4.2   Perceptual and Cognitive Pressures in Melodic Evolution

The compression of interval size and the regular melodic surface found in our data obey two basic laws of perceptual organization: proximity and good continuation [17]. These principles entail that the auditory system creates pattern units using elements that are close in pitch and ordered in a single melodic direction, whereas boundaries are introduced at points where changes in the interval or direction occur. In general, tone streams governed by those principles produce perceptual cohesion and can be efficiently encoded in memory. A different Gestalt construct might explain the emergence of melodic transformations within a melodic pool, such as the perceived equivalence for

listeners of pairs of melodic sequences with a symmetry relation, especially inversions [19, 61]. By using high-order abstractions of a given melodic sequence, such as transforms, participants introduced moderate levels of redundancy in a code, thus reducing memory load while maintaining melodic diversity and expressivity. Overall, our results are supported by psychological studies showing that melodic information is processed more effectively when it follows principles of auditory grouping (e.g., [58, 59]).

These results suggest that melodic information is recalled as particular instances of grouping or specific configurations, but they also provide indications as to how these configurations may be encoded and recalled. We noticed that fine-grained information was usually lost, whereas the contour of melodies was more often encoded [21]. As a result, we found that code transmission was enhanced, as revealed by melodic contour changes rather than changes in pitch sequences. Similar results were originally reported by Bartlett and Burt [3] in serial production experiments. In that study, the oral tradition of a folk tale was reproduced at a smaller scale. The authors discovered that only "dominant" features were well remembered in retellings, while details were omitted or modified. Based on this result, they proposed that remembering is more a reconstructive process based on a high-level abstract model than a replication of sensory information. This hypothesis is now supported by empirical work in language acquisition. According to this view, the linguistic input must be rapidly re-coded in compressed forms through multilevel chunking mechanisms to avoid it being overridden by incoming information. Following this process most details are lost, whereas the new memory represents just an abstract summary of the original sensory input [12]. Results in our experiment suggest that similar chunk-and-pass mechanisms may operate in the evolution of melodic material. These mechanisms may affect the transmission and evolution of melodic structures, historically documented in the musical domain [39, 60]. The qualitative observations by Verhoef [79] are in line with our findings.

The effects we have observed correspond to widespread trends or melodic regularities found across the world, with some overlap between the musical and the speech domain. The arrangement of sounds in intervals of smaller size (proximity) [32], with coherent directions, is a well-documented phenomenon in different musical [31] and linguistic cultures, when continuous pitch glides are converted into discrete pitch patterns [53]. The "greater emphasis on global features than on local details" [78, p. 427], and likewise the proportion of transformations found in our experiment, are also found in different cultures and historical periods [19, 31].

Studies using iterated learning show that cultural systems, when they are acquired and reused, are regularized so that structure becomes *compressible*, or easier to acquire or reproduce. Compression is a fundamental cognitive principle [57]. It allows for the most concise encoding of a serial input, and as a result it may enhance learnability or ease the burden on memory storage [10]. When this principle is applied to languages, and when expressiveness pressures also come into play, it is partly reflected in a key feature of (certain fragments of) natural language and thought: *compositionality*. The systematic arrangement of signal elements in relation to the meanings they express provides a parsimonious encoding of meaning and an economical use of expressions in the language [35, 79]. The negligible changes in the degree of compositionality found in this experiment stand in contrast with previous iterated learning studies [35, 75, 38], and are instead consistent with an arrangement of patterns of sounds based on perceptual principles [24, 55]. In support of this conclusion, a relatively inefficient transmission of compositional structures was reported in musically naive participants for a miniature artificial tone language [70]. We cannot exclude that the present outcome might change if a full compositional semantics, as in Kirby et al.'s experiments [35], were made available to participants. However, precisely because a trivial compositional solution could be achieved in our experiment, we may interpret our result as preliminary evidence that participants were not restructuring the codes following strictly linguistic principles.

In parallel, we observed a progressive compression of melodic material, as indicated by an increase in the similarity between signals of the same melodic set. At first, this might be interpreted as a progressive evolution towards monomelodic systems. The tendency of signaling systems towards homonymy has been reported in previous iterated learning studies [35, 79]. These findings are in line

with the notion that memory pressure (one form of *cognitive economy*) is at work in the cultural evolution of communicative systems [10, 71], leading towards greater compressibility within the system in use. Several effects of these pressures are well attested in human musical and vocal behavior. One relevant example is the combinatorial use of a limited set of melodic segments [1, 28]. However, our results are also compatible with an alternative explanation: the progressive combinatorial use of a limited set of melodic segments. In the last generations, fewer melodic segments are reused in combination to produce the full set of signals. Hence, the similarity between signals is expected to increase. In the musical domain, this is reflected in the preferred use across traditions of small systems of discrete pitch elements [5–7], and in the combinatorial reuse of these elements, or groups of elements (motives [2]), to produce a virtually boundless number of complex musical phrases [79]. A pressure for the system to become expressive, as in our experiment, would point towards this direction. It is worth noting that both explanations refer to adaptations to the limits of human memory, and they are not necessarily mutually exclusive.

In sum, our results are in contrast with the notion that principles of *compressibility*, primarily *compositionality*, play a prominent role in the evolution of artificial symbolic systems. If constraints like compressibility were the main source of change, we should observe a gradual *increase* in monotones [44], with compositionality also emerging. Iterated learning experiments using non-auditory (e.g., visual) signals support in part this notion [13]. However, we did not observe such trends. Compositionality was low, and monotones represented only a small fraction of the entire pool of signals. While we saw the emergence of regularities over multiple generations, those seem to respond more to principles for perceiving, remembering, and producing tone sequences [18] than for organizing language-like symbol sequences [57]. Critically, unlike language properties such as compositionality, such principles of organization may be difficult to explain in the classical framework of information theory.

5   A Broader View of Cultural Evolution

One key tenet in cultural evolution research is that information, once acquired from others, is transformed by individuals in some nonrandom manner to fit functional and structural constraints of their mind/brain [5]. These constraints are amplified during cultural transmission to produce large-scale population-level patterns [33]. This may account for universal aspects of culture [11, 16]. Data in support of this notion derive from modeling work, where cultural evolution has been simulated on a small scale in artificial conditions. Most of these studies include agent-based modeling and human experiments that implement diffusion-chain methods [6, 34, 35, 38, 67, 80].

The reliability of the data produced by cultural evolution models [56] has opened new lines of research in several fields of inquiry, from social psychology to language evolution [49]. Omissions and modifications of cultural material, as observed in human diffusion chains, are taken to reveal the effects of cognitive constraints or inductive biases on cultural evolution [50]. In this regard, ease of processing proved to be a major force in the emergence of population-level cultural patterns [49, 27]. On the other hand, the use of diffusion chains in computer simulations and mathematical analyses has proved useful to produce a formal characterization of the mechanisms and the effects of constraints on cultural transmission (e.g., [27, 33]). Experimental, computational, and analytic approaches, however, converge on similar conclusions: in the long run, signaling systems increasingly reflect processing biases and constraints of individuals (e.g., [27]), so that the forms that are easier to learn and process become the most prevalent in the population [36]. One possible criticism of the hypothesis of a cognitive foundation of culture [68] is that the existence of these constraints has so far been only inferred from behavioral data. However, in a recent experiment [46], we provided the first neurophysiological evidence in support of this view, showing that the emergence of melodic regularities in the course of cultural transmission (precisely, proximity and good continuation), is driven by the information-processing capacity of individuals.

Here, we show that the effects of cognitive constraints on melodic processing [46] are magnified during transmission in diffusion chains, reproducing some universal aspects of musical systems. This result is, in this regard, in line with previous cultural evolution studies, including language [35, 83]. It provides initial empirical evidence for a view of music as an evolutionary adaptive system [77], as has been proposed for language [14]. At the same time, our work suggests that partly different principles drive the evolution of these two symbolic systems and, most probably, cultural symbolic systems in general. During transmission, tone sequences were mainly shaped by perceptual principles of auditory sequence organization, rather than by strict economy principles, which instead seem to provide a major force of change in languages [35]. This issue will have to be addressed in controlled experimental settings in future research.

6   Conclusion

Our results suggest that culturally transmitted artificial tone sequences are subject to the same set of basic cognitive and perceptual constraints that operate on the cultural evolution of auditory symbolic systems [7], musical and speech melodies included [77]. When these principles are brought out through several cycles of learning and use, they can explain the emergence of common melodic properties in auditory symbolic systems. The present work is the first experimental contribution to debate on the role of cognitive constraints in the cultural evolution of tone systems.

References

1. Adams, C. R. (1976). Melodic contour typology. *Ethnomusicology*, *20*(2), 179–215.

2. Bartel, C. (2006). Musical thought and compositionality. *Postgraduate Journal of Aesthetics*, *3*(1), 25–36.

3. Bartlett, F. C., & Burt, C. (1933). Remembering: A study in experimental and social psychology. *British Journal of Educational Psychology*, *3*(2), 187–192.

4. Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes.

5. Boyd, R., & Richerson, P. J. (1988). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

6. Brighton, H., & Kirby, S. (2006). Understanding linguistic evolution by visualizing the emergence of topographic mappings. *Artificial Life*, *12*(2), 229–242.

7. Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

8. Caldwell, C. A., & Millen, A. E. (2008). Studying cumulative cultural evolution in the laboratory. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1509), 3529–3539.

9. Caldwell, C. A., & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PLOS ONE*, *7*(8), e43807.

10. Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22.

11. Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(05), 489–509.

12. Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, 1–52.

13. Cornish, H., Smith, K., & Kirby, S. (2013). Systems from sequences: An iterated learning account of the emergence of systematic structure in a non-linguistic task. In M. Knauff (Ed.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 340–345). New York: Curran Associates.

14. Cornish, H., Tamariz, M., & Kirby, S. (2009). Complex adaptive systems and the origins of adaptive structure: What experiments can tell us. *Language Learning*, *59*, 187–205.

15. Davies, S. (1994). *Musical meaning and expression*. Ithaca, London: Cornell University Press.

16. Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain*. New York: Norton.

17. Deutsch, D. (2013). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (pp. 183–248). New York: Academic Press.

18. Deutsch, D. (2013). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (pp. 249–325). New York: Academic Press.

19. Dowling, W. J. (1972). Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics*, *12*(5), 417–421.

20. Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, *49*(2B), 524–531.

21. Dowling, W. J. (1994). Melodic contour in hearing and remembering melodies. In R. Aiello & J. Sloboda (Eds.), *Musical perceptions* (pp. 173–190). New York: Oxford University Press.

22. Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the U.S.A.*, *111*(15), E1454–E1462.

23. Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial clues*. Los Altos, CA: Malor Books.

24. Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*(1), 173–215.

25. Fitch, W. (2005). The evolution of music in comparative perspective. *Annals of the New York Academy of Sciences*, *1060*(1), 29–49.

26. Flynn, E., & Whiten, A. (2008). Cultural transmission of tool use in young children: A diffusion chain study. *Social Development*, *17*(3), 699–718.

27. Griffiths, T. L., Kalish, M. L., & Lewandowsky, S. (2008). Review. Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1509), 3503–3514.

28. Halliday, M. A. K. (1981). *A course in spoken English: Intonation*. London: Oxford University Press.

29. Hamming, R. W. (1986). *Coding and information theory*. Upper Saddle River, NJ: Prentice-Hall.

30. Hsieh, S., Hornberger, M., Piguet, O., & Hodges, J. R. (2012). Brain correlates of musical and facial emotion recognition: Evidence from the dementias. *Neuropsychologia*, *50*(8), 1814–1822.

31. Huron, D. (1996). The melodic arch in Western folksongs. *Computing in Musicology*, *10*, 3–23.

32. Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception: An Interdisciplinary Journal*, *19*(1), 1–64.

33. Kirby, S. (2001). Spontaneous evolution of linguistic structure—an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102–110.

34. Kirby, S. (2002). Natural language from artificial life. *Artificial Life*, *8*(2), 185–215.

35. Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the U.S.A.*, *105*(31), 10681–10686.

36. Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114.

37. Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. C. Laurea & M. A. Domenico Parisi (Eds.), *Simulating the evolution of language* (pp. 121–147). New York: Springer Verlag.

38. Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

39. Kleeman, J. E. (1985). The parameters of musical transmission. *The Journal of Musicology*, *4*(1), 1–22.

40. Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, *8*(2), 89–105.

41. Koopman, C., & Davies, S. (2001). Musical meaning in a broader perspective. *The Journal of Aesthetics and Art Criticism*, *59*(3), 261–273.

42. Leman, M. (1992). The theory of tone semantics: Concept, foundation, and application. *Minds and Machines*, *2*(4), 345–363.

43. Lense, M. D., Gordon, R. L., Key, A. P., & Dykens, E. M. (2014). Neural correlates of cross-modal affective priming by music in Williams syndrome. *Social Cognitive and Affective Neuroscience*, *9*(4), 529–537.

44. Leroi, A. M., MacCallum, R. M., Mauch, M., & Burt, A. (2012). Reply to Claidière et al.: Role of psychological bias in evolution depends on the kind of culture. *Proceedings of the National Academy of Sciences of the U.S.A.*, *109*(51), E3527.

45. Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

46. Lumaca, M., & Baggio, G. (2016). Brain potentials predict learning, transmission and modification of an artificial symbolic system. *Social Cognitive and Affective Neuroscience*, *11*(12), 1970–1979.

47. Mathews, M. V., Pierce, J. R., Reeves, A., & Roberts, L. A. (1988). Theoretical and experimental explorations of the Bohlen–Pierce scale. *The Journal of the Acoustical Society of America*, *84*(4), 1214–1222.

48. Mertens, P. (2004). The prosogram: Semi-automatic transcription of prosody based on a tonal perception model. In *Speech Prosody 2004, International Conference*.

49. Mesoudi, A. (2015). Cultural evolution: A review of theory, findings and controversies. *Evolutionary Biology*, 1–17.

50. Mesoudi, A., & Whiten, A. (2008). Review. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1509), 3489–3501.

51. Moreno, M., & Baggio, G. (2015). Role asymmetry and code transmission in signaling games: An experimental and computational investigation. *Cognitive Science*, *39*(5), 918–943.

52. Nowak, I., & Baggio, G. (2016). The emergence of word order and morphology in compositional languages via multigenerational signaling games. *Journal of Language Evolution*, *1*(2), 137–150.

53. Patel, A. D. (2010). *Music, language, and the brain*. New York: Oxford University Press.

54. Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences of the U.S.A.*, *110*(22), 8836–8841.

55. Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

56. Richerson, P. J., & Boyd, R. (1985). Simple models of complex phenomena: The case of cultural evolution. In J. Dupre (Ed.), *The latest on the best: Essays on evolution and optimality* (pp. 27–52). Cambridge, MA: MIT Press.

57. Rissanen, J. (1978). Modeling by shortest data description. *Automatica: The Journal of IFAC, The International Federation of Automatic Control*, *14*(5), 465–471.

58. Rohrmeier, M., & Cross, I. (2013). Artificial grammar learning of melody is constrained by melodic inconsistency: Narmour's principles affect melodic learning. *PLOS ONE*, *8*(7), e66174.

59. Rohrmeier, M., Rebuschat, P., & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness and Cognition*, *20*(2), 214–222.

60. Sawa, G. D. (2002). Oral transmission in Arabic music, past and present. *Oral Tradition*, *4*(1), 254–265.

61. Schoenberg, A., & Newlin, D. (1951). *Style and idea*. London: Williams and Norgate.

62. Schönpflug, U. (2008). *Cultural transmission: Psychological, developmental, social, and methodological aspects*. Cambridge, UK: Cambridge University Press.

63. Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, *14*(9), 411–417.

64. Shannon, C. (1949). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.

65. Siegel, S. (1988). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

66. Skyrms, B. (2010). *Signals: Evolution, learning, and information*. Oxford, UK: Oxford University Press.

67. Smith, K., Kirby, S., & Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, *9*(4), 371–386.

68. Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Cambridge, UK: Cambridge University Press.

69. Swarup, S., & Gasser, L. (2009). The iterated classification game: A new model of the cultural transmission of language. *Adaptive Behavior*, *17*(3), 213–235.

70. Tamariz, M., Brown, J. E., & Murray, K. M. (2010). The role of practice and literacy in the evolution of linguistic structure. In A. D. M. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *The evolution of language: Proceedings of the 8th International Conference*. Singapore: World Scientific.

71. Tamariz, M., & Kirby, S. (2015). Culture: Copying, compression, and conventionality. *Cognitive Science*, *39*(1), 171–183.

72. Tamariz, M. (2013). RegMap (Version 1.0) (software).

73. Tan, R., & Fay, N. (2011). Cultural transmission in the laboratory: Agent interaction improves the intergenerational transfer of information. *Evolution and Human Behavior*, *32*(6), 399–406.

74. Terhardt, E. (1987). Gestalt principles and music perception. In W. A. Yost & C. S. Watson (Eds.), *Auditory processing of complex sounds* (pp. 157–166). Hillsdale, NJ: Erlbaum.

75. Theisen-White, C., Kirby, S., & Oberlander, J. (2011). Integrating the horizontal and vertical cultural transmission of novel communication systems. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 956–961). Austin, TX: Cognitive Science.

76. Tomasello, M. (2009). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

77. Trainor, L. J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philosophical Transactions of the Royal Society B*, *370*(1664), 20140089.

78. Trehub, S. E. (2000). Human processing predispositions and musical universals. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 427–448). Cambridge, MA: MIT Press.

79. Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, *4*(4), 357–380.

80. Wheeler, M., Bullock, S., Di Paolo, E., Noble, J., Bedau, M., Husbands, P., & Seth, A. (2002). The view from elsewhere: Perspectives on ALife modeling. *Artificial Life*, *8*(1), 87–100.

81. Whiten, A., Spiteri, A., Horner, V., Bonnie, K. E., Lambeth, S. P., Schapiro, S. J., & De Waal, F. B. (2007). Transmission of multiple traditions within and between chimpanzee groups. *Current Biology*, *17*(12), 1038–1043.

82. Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth curve analysis and generalized additive modeling. *Journal of Language Evolution*, *1*(1), 7–18.

83. Xu, J., Dowman, M., & Griffiths, T. L. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1758), 20123073.