Music adapting to the brain:

From diffusion chains to neurophysiology

The research presented in this thesis was carried out at the Neuroscience Group at the International School for Advanced Studies in Trieste, and at the Center for Music in the Brain, University of Aarhus.

Lay-out, figures: Massimo Lumaca, Elena Kappers
Cover: Elena Kappers, Cristina De Silvestris
Drawings: Cristina De Silvestris
Photographer: Eleonora Cugini

International School for Advanced Studies
Trieste

Music adapting to the brain:

From diffusion chains to neurophysiology

Doctoral thesis

The public defense will take place on
9 February 2017 at 13:30

Candidate:

Massimo Lumaca

born on 27 April 1984
in Rome

Supervisor

Dr. Giosuè Baggio

Department of Language and Literature,
Norwegian University of Science and Technology, Norway

Examiners

Dr. Bruno Gingras

Department of Psychology,
Universität Innsbruck, Austria

Dr. Mònica Tamariz

Department of Linguistics and English Language,
The University of Edinburgh, Scotland

Dr. Davide Crepaldi

Neuroscience Area,
International School for Advanced Studies, Italy

Dr. Domenica Bueti

Neuroscience Area,
International School for Advanced Studies, Italy

Alla mia famiglia

Abstract

During the last decade, the use of experimental approaches on cultural evolution research has provided novel insights, and supported theoretical predictions, on the principles driving the evolution of human cultural systems. Laboratory simulations of language evolution showed how general-domain constraints on learning, in addition to pressures for language to be expressive, may be responsible for the emergence of linguistic structure. Languages change when culturally transmitted, adapting to fit, among all, the cognitive abilities of their users. As a result, they become regular and compressed, easier to acquire and reproduce. Although a similar theory has been recently extended to the musical domain, the empirical investigation in this field is still scarce. In addition, no study to our knowledge directly addressed the role of cognitive constraints in cultural transmission with neurophysiological investigation.

In my thesis I addressed both these issues with a combination of behavioral and neurophysiological methods, in three experimental studies. In **study 1** (Chapter 2), I examined the evolution of structural regularities in artificial melodic systems while they were being transmitted across individuals via coordination and alignment. To this purpose I used a new laboratory model of music transmission: the multi-generational signaling games (MGSGs), a variant of the signaling games. This model combines classical aspects of lab-based semiotic models of communication, coordination and interaction (horizontal transmission), with the vertical transmission across generations of the iterated learning model (vertical transmission). Here, two-person signaling games are organized in diffusion chains of several individuals (generations). In each game, the two players (a sender and a receiver) must agree on a common code - here a miniature system where melodic riffs refer to emotions. The receiver in one game becomes the sender in the next game, possibly retransmitting the code previously learned to another generation of participants, and so on to complete the diffusion chain. I observed the gradual evolution of several structures features of musical phrases over generations: proximity, continuity, symmetry, and melodic compression. Crucially, these features are found in most of musical cultures of the world. I argue that we tapped into universal processing mechanisms of structured sequence processing, possibly at work in the evolution of real music. In **study 2** (Chapter 3), I explored the link between cultural adaptation and neural information processing. To this purpose, I combined behavioral and EEG study on 2 successive days. I show that the latency of the mismatch negativity

(MMN) recorded in a pre-attentive auditory sequence processing task on day 1, predicts how well participants learn and transmit an artificial tone system with affective semantics in two signaling games on day 2. Notably, MMN latencies also predict which structural changes are introduced by participants into the artificial tone system. In **study 3** (Chapter 4), I replicated and extended behavioral and neurophysiological findings on the temporal domain of music, with two independent experiments. In the first experiment, I used MGSGs as a laboratory model of cultural evolution of rhythmic equitone patterns referring to distinct emotions. As a result of transmission, rhythms developed a universal property of music structure, namely temporal regularity (or isochronicity). In the second experiment, I anchored this result with neural predictors. I showed that neural information processing capabilities of individuals, as measured with the MMN on day 1, can predict learning, transmission, and regularization of rhythmic patterns in signaling games on day 2. In agreement with study 2, I observe that MMN brain timing may reflect the efficiency of sensory systems to process auditory patterns. Functional differences in those systems, across individuals, may produce a different sensitivity to pressures for regularities in the cultural system. Finally, I argue that neural variability can be an important source of variability of cultural traits in a population.

My work is the first to systematically describe the emergence of structural properties of melodic and rhythmic systems in the laboratory, using an explicit game-theoretic model of cultural transmission in which agents freely interact and exchange information. Critically, it provides the first demonstration that social learning, transmission, and cultural adaptation are constrained and driven by individual differences in the functional organization of sensory systems.

Acknowledgments

When I started the master in Neurobiology, I had a dream: to combine my personal interest for cognitive neuroscience and human behavior, with the greatest passion of my life, namely music. In SISSA I had the opportunity to make this dream true. During these 4 years I have met many people that have been important to me, in many aspects. Here, I'd like to thank each of them.

Firstly, I am really grateful to my supervisor **Giosuè**. He was the first to believe in me when I tried the PhD entrance the first time, but I wasn't successful. He accepted me as predoc student in his laboratory, giving me a second chance and sharing a believe in my dream. I can't thank him enough for that, because now I am here defending the doctorate. The work has been hard since he left SISSA. Nonetheless, he was always present, and taught me so much about how to conduct good research. He formed me as an independent mind, teaching me to get by on my own when faced with novel problems, which is a requested skill for who wants to undertake the difficult path of research. I cannot be more grateful to him for that. Thank you!

I want to thank my family, my mum **Cecilia**, my father **Lamberto**, and my sister **Francesca.** They supported me all these years to achieve the goals I had set out, musical and academics, in ups and downs. We went through a bad moment few years ago, but now we are all together, which is what really matters. You are a wonderful family, I couldn't ask more! I also want to thank my **uncles** (zio Vincenzo and zio Massimo), **aunts** (zia Nella and zia Loredana), and **cousins** (Ale, Cricri, Lalla, and Sara) which followed my adventure from far, but they were always present for me. My family now includes also other people, **Anna**, **Bert**, and **Martin**, which I also want to thank for their continuous support, warmth and help (practical and motivational). It is thanks to them that I noticed the vacant SISSA fellowship, and my adventure began. I wouldn't be here without them. Thank you very much!

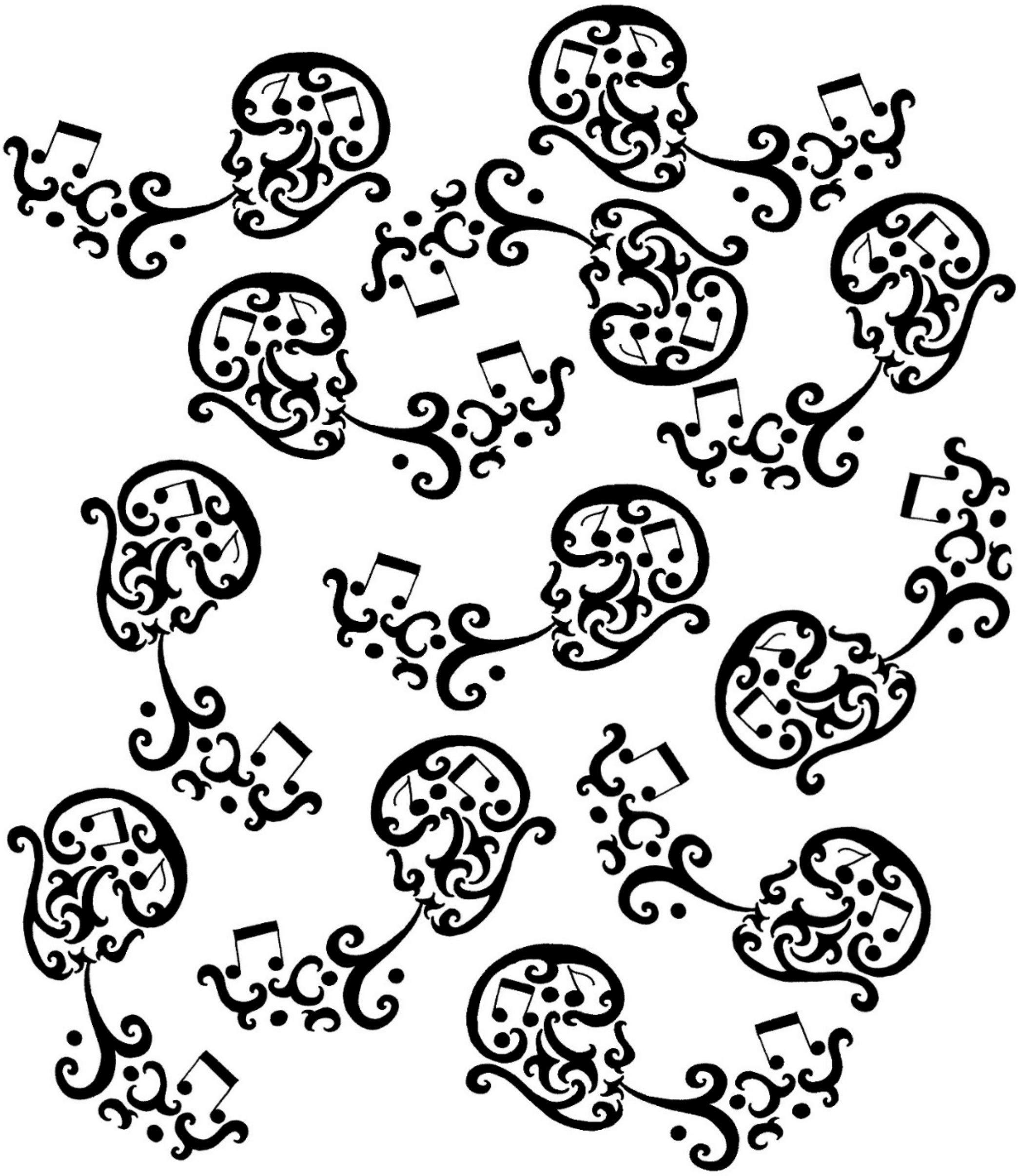
I have no words to thank enough my girlfriend **Elena**. She made me enjoy this experience, standing close to me in the most difficult moments, despite the fact that she lives in another country. Sometimes she also helped me to conduct my experimental research, preparing the subjects during the EEG (fun!), or managing the organization of participants during the transmission chains. We went through a lot during these years, but we shared beautiful moments: I am lucky to have someone like you. Thank you!

I would especially like to thank my closest friends in SISSA, **Carlotta**, **Cinzia**, **Luca**, **Vanni**, **Fabrizio**, **Sebastian**, and **Carol**, for the experiences, often funny, we shared together in (and outside) the School. You have been for me an anchor during these 4 years, I really hope we will be still in touch at the end of this experience. Let me say this: you are great people! I also want to thank my other friends **Lorenzo**, **Mattia**, **Antonio**, **Richard**, **Cristiano, Daniele**, **Ludovica, Lara**, **Alessio**, **Iga** and **Ana Laura**, **Annabel** and **Gabriela,** and all the people I had the pleasure to play the piano with at Caffè San Marco.

Finally, I am grateful to the people of the Music in the Brain (MIB), which made my ERASMUS experience in Aarhus beautiful and productive: **Elvira**, **Peter**, **Niels**, **Maria Celeste**, **Maria**, **Henrique**, **Patrizia**, **Stine**, **Leonardo** and **Alessandro**. I will spend there my next year as a postdoc: I hope to enjoy this experience as much as I did in the past.

Thank you all!

Contents



Chapter 1

General introduction: Music evolution in the laboratory

Massimo Lumaca

General introduction

Music is a universal trait of human culture with an ancient history. It can be found in every known human society, past and present. It shows a remarkable diversity across world's cultures, but also deep analogies. It is not surprising then that it has acquired a particular charm for biological and cognitive sciences. Why is music organized like this, and not some other way? How can we account for both variable and invariant aspects of musical structure?

Theoretical and methodological advances in the field of language evolution can be of help to shed light on this complex matter. Over the past decade, the mechanisms underlying the cultural evolution of languages have been extensively studied. These works highlighted the role of cultural transmission as a powerful driving force in the evolution of natural languages. It is hypothesized that languages adapt to constraints of the human brain/mind when they are culturally transmitted, developing features that make them easier to acquire and transmit. This view is gaining some empirical support from agent-based simulations, mathematical modelling, and studies of cultural transmission in the laboratory, where it has been invoked to account for universal aspects of human language. Although similar theories having been extended to the musical domain, empirical investigation in this field is still scarce. Furthermore, no study to our knowledge addressed with neurophysiological investigation the question whether cultural information reflect human brain information processing.

My thesis has two main aims. First, I want to provide evidence that some of the *universal* properties found in musical structure can be directly explained as adaptations of 'musical systems' (in fact artificial systems of simple tone sequences) to perceptual and cognitive constraints[1] of the human brain/mind during cultural transmission. Our second objective is to provide an initial neural evidence to this phenomenon. In addition, we aim to relate the kind of 'neural diversity' found across individuals to some forms of *variation* found in musical behavior. I will address these questions by combining experimental methods of evolutionary linguistics and cognitive neuroscience. In one set of studies I observe how artificial (auditory) symbolic systems change when transmitted from senders to receivers thorough coordination and communication, along diffusion chains in which the receiver in one game becomes the sender in the next game. Then, independently, I examine whether individual variation in neural information processing, as revealed by ERPs, predicts variation in cultural behavior: specifically, how well participants learn a signaling system, how faithfully retransmit it, and which

---

[1] Constraints on information processing and memory capacity. We originally decided to adopt this term instead of the more used *cognitive biases*, to avoid any confusion with judgment biases of social psychology, the latter open to individual awareness

modifications are introduced. I do this separately for two basic dimensions of musical structure, namely *rhythm* and *melody*.

This thesis focuses on the *cultural* evolution of musical structure, and its neural underpinnings. Despite the fact that some of the findings of this work might be explanatory for the cultural evolution of symbolic systems in general, results will be specifically interpreted on the domain of music.

1.1. The origins of musical structure: a brief historical overview

There has recently been a surge of interest in the origins and the evolution of music (McDermott & Hauser, 2005; Miller, Wallin, Merker, & Brown, 2000; Patel, 2010). Among the many questions concerning its biological and cognitive foundations (Honing, ten Cate, Peretz, & Trehub, 2015), that of the origins of structure in music is one of the most controversial. How does structure arise in music? How do we explain the vast diversity in musical variants exhibited among (and within) the musical traditions of the world (Lomax, 1977, 1980)? And how to account for the evidence that few of these variants are present in nearly all musical cultures, with frequencies significantly above the chance level (Savage, Brown, Sakai, & Currie, 2015)? We believe that neural constraints on cultural transmission can be partially responsible for both aspects. Here I provide a brief summary of the barriers that systematic musicology had to face in the last century addressing some of these questions, and also potential solutions from the language evolutionary field.

Music is unique among the biological sound systems observed in nature, and can be considered with language a prominent and distinctive characteristic of humankind (Patel, 2010). In particular, music reveals some distinctive features that, taken together, cannot be found either in language or in any non-human vocalization system (Fitch, 2006).

Addressing how these features evolved in music is far from a trivial task. First, music, unlike other biological and cultural artefacts, did not leave accurate information about its past forms. Audio recordings and written notation are only relatively recent phenomena when compared to the putative origins of music (Conard, Malina, & Münzel, 2009; Fitch, 2006). The information we can extrapolate from naturalistic data is far too small to be sufficiently valuable.

Second, the nature of music, cultural or biological, is still debated. The predominant position of the last century views music as a fully socio-cultural construct, free to vary and with virtually no biological constraints on its design features (the '*cultural account*') (Blacking, Byron, & Nettl, 1995; Nettl, 1983). Based on this notion, humans can appreciate and acquire any musical language, if exposed to it for a sufficient time (Schoenberg & Newlin, 1951). Many, if not all properties of musical systems would be thus an arbitrary results of history; citing a quote by Lewontin (2000) “[...] wandering phenomena through the space of possibilities without any law-like generalities”. The vast diversity of musical forms, observed across cultures and over historical time, supports this idea.

This position has been recently challenged by the steady, increasing evidence by psychological and neuroscientific methods of biological constraints in music perception and memory (the ‘*biological account*’). Recent experimental work reveal common patterns of pitch and temporal processing in humans, with parallels in the animal kingdom (Fitch, 2015). These mechanisms emerge early in development (Trehub, 2000; Winkler, Háden, Ladinig, Sziller, & Honing, 2009), irrespective of individual’s musical experience and culture (Drake & Bertrand, 2001; Stevens & Byron, 2009). Following this view, prototype properties of music, such as a relatively steady beat (termed ‘isochronicity’), and a small distance between adjacent tones (or ‘pitch proximity’), are granted and constrained by built-in functional properties of the brain (Jackendoff & Lerdahl, 1983), which manifest naturally in human cultures (Lerdahl, 1992). This can potentially explain the wide geographic distribution, and the greater than chance global frequencies recently observed for some of these properties by phylogenetic studies (Savage, Brown, Sakai, & Currie, 2015). Whether these built-in mechanisms are domain-specific modules (Huron, 2012; Peretz & Coltheart, 2003), and, if so, whether they are product of evolutionary pressures, sexual or natural selection, is still in the field of debates, or even more, of speculations (for reviews see Justus & Hutsler, 2005; McDermott & Hauser, 2005).

An important point to be made here is that neither of these two accounts, taken separately, can provide a completely satisfying explanation for a comprehensive theory of the origins and the evolution of musical structure. On one hand, music is not an arbitrary cultural construct: it firmly rests on biological foundations, such as memory capacity, hierarchical processing, and auditory scene analysis (ASA) mechanisms (Deutsch, 2012; Snyder, 2008). This is now a core tenet in the fields of music cognition and psychology. However, how this biological endowment would manifest itself from single individuals to musical structure at population-scale is less clear: an issue already known in linguistics as the ‘*problem of linkage*’ (Kirby, 1999).

Critically, these seemingly opposite views can be reconciled, and the issue of linkage solved, if we think of music neither as a uniquely cultural or biological artefact, but as a cultural adaptive system that has been shaped for many thousands of years by cycles of learning, use and transmission (Morley, 2013). In this view different aspects of the temporal (rhythm and meter) and spectral (pitch) structure of music, result as adaptations to a combination of cognitive constraints and socio-cultural forces (Merker, 2006; Merker, Morley, & Zuidema, 2015; Trainor, 2015). This idea was originally developed in the field of evolutionary linguistics to explain the emergence of structure in human languages, and to account for (near-)universal aspects of linguistic structure (Christiansen & Chater, 2008; Deacon, 1997). Following this view, inter-generational transmission exerts profound effects on the form and the structure of language (Kirby, 1999). Like language, music is a rich syntactic system of perceptually discrete elements, which can be found in all human cultures, and which is culturally transmitted (Fitch, 2006; Patel, 2010). Given the parallels between language and music, it seems useful to extend the same idea to the musical domain. In the following sections I examine this idea in detail and describe an empirical method aimed to test it.

1.2. Music as a complex adaptive system

Any form of knowledge that is manifested in behavior and preserved over time through social learning mechanisms can be regarded as ‘cultural’ (Boyd & Richerson, 2005). One common misconception is that evolution belongs uniquely to biological organisms, while it is used metaphorically when referring to culture. However, when Darwin formulated his theory on evolution, he did not know about genes. In addition, he first noticed a parallel between the evolution of species and that of languages (Darwin, 1888). Darwin noticed that despite the striking variability, languages show similar characteristics like species do. In addition, he observed that languages compete among themselves: dominant languages (or also tiny linguistic variants) can drive others to extinction.

When the mechanisms underlying the transmission and evolution of culture began to be investigated more formally (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981) this misconception was set aside. In its broad sense, the term evolution simply indicates a change over time in the frequency of traits (or ‘variants’) in the population. Briefly, this change stands on three main principles: *variation, inheritance* and *selection*. The principles hold for all forms of evolution, including cultural (Mesoudi, Whiten, & Laland, 2004). An evolutionary system must show (*i*) variation in its traits, (*ii*) transmission of variants, (*iii*) competition between transmitted variants to be acquired and expressed. Any system of knowledge that fits these three criteria is a cultural system.

Music is maintained over time through different forms of social transmission. Besides being culturally transmitted, it shows an incredible variety in its traits (within and between cultures) (Rzeszutek, Savage, & Brown, 2012), and some of its variants have larger chance to be transmitted than others. Musical change can be thus considered a cultural evolutionary process. In addition, music can be viewed as a system of complex and interdependent organisms, the musical features, which evolve under the selective pressures of the human brain. In other words music, like language (Cornish, Tamariz, & Kirby, 2009), seems to fit most of the criteria to be considered a cultural adaptive system on its own right, and *cultural evolution* is the fundamental mechanism that makes this possible.

1.3. Cultural transmission solves the problem of linkage

Music, like language, is the product of thousands of years of transmission via social mechanisms (teaching, imitation, and observation), and innovations, introduced by single individuals via asocial mechanisms (e.g., musical compositions, embellishments, etc.). This thesis focuses on the first class of mechanisms, and in particular on the oral transmission of music, which constitutes, still today, the principal mechanism among world cultures of transmitting music (see section 1.7.1.1). In oral traditions, there is no musical notation to relieve the individuals of memory and perceptual loads. An individual can only rely on his brain, with all its limits, as the repository of cultural information and as the vector of transmission. This way of music acquisition (and transmission), provides insights on the evolutionary pressures that may have produced some of its fundamental characteristics. Above all, if music wants to survive the intergenerational transfer it must be learnable.

When passed on from from one musician to another, musical culture has first to survive to the *memory bottleneck* (Deacon, 1997), the limited capacities of the receivers to induce (and store) the rules that organize the musical culture, or ‘musical grammars’. Cognitive and perceptual limitations on grouping (Bregman, 1990; Jackendoff & Lerdahl, 1983), capacity and temporal limits of auditory memory (Bottcher□Gandor & Ullsperger, 1992), serial processing (Sigman & Dehaene, 2005), attention (Cohen, 2014), and neurodynamics of the auditory nervous system (Large, 2010), are probably only some of the properties and constraints that constitute the memory bottleneck. In addition, the receiver is only exposed to a portion of music of its own culture (or *information bottleneck*) (Kirby & Hurford, 2002). Limitations on cognitive capabilities and limited access to data considerably restrict the properties of the musical lore that will survive to the next generation.

These limitations have major implications for the evolution of music. In each generational transfer, traits of the stimulus transmitted will be in competition with each other to be encoded sufficiently fast in the learner’s memory. Noise and new-coming information make this competition even harder, overwriting or interfering with the traits which still must be encoded (Christiansen & Chater, 2016). Only the fittest (i.e., the easier and faster to process) will be encoded in the learner’s memory. Accordingly, sustained filtering processes like this and even minimal differences in processability will be reflected in the musical corpus (Kirby, Dowman, & Griffith, 2007). Major properties of musical structure can be thus understood as cultural adaptations to sustained filtering processes by human brain throughout cultural history (Merker et al., 2015). Cultural transmission and learning bottleneck provide an elegant solution to the problem of *how* universal learning mechanisms of single individuals can manifest as invariant aspects of musical structure at population-level (Merker et al., 2015; Trehub, 2015).

The solution to the ‘problem of linkage’ was first formulated in the field of evolutionary linguistics (Deacon, 1997). In the last decade, this view was empirically tested by modelling cultural evolution of language at small-scale using *linear diffusion chain methods* (Bartlett, 1932). In the next section I provide a brief overview of the most prominent of these methods, with relative applications on language and music evolution.

1.4. Iterated learning model

Historical sciences such as linguistics or musicology aim to describe human symbolic systems at three main levels of adaptation: phylogeny, ontogeny and glossogeny (Hurford, 1999). If the first level describes the *biological* adaptation of the learning machinery underlying cultural transmission (Fitch, 2010; Honing et al., 2015), the other two are concerned with *cultural* adaptations that occur in music and language at individual- and population-scale respectively (Christiansen & Kirby, 2003; Fitch, 2015). Specifically, ontogeny refers to the development of the musical or linguistic competence in the lifetime of an individual. Glossogeny refers instead to changes observed in cultural systems in historical eras, mostly result of cumulative deviations from the original model.

Glossogeny in particular received a great deal of attention in the past decade, being virtually explanatory for the emergence of population-level design features in these systems. This is the domain where models of cultural transmission operate and test hypotheses on cultural evolution; the most prominent of these models is the *iterated learning model* (or ILM) (Kirby, 2001). Over the past two decades, this model provided the main empirical framework for understanding the mechanisms of cultural emergence, transmission and evolution of human symbolic systems. Quoting the definition of iterated learning by Scott-Phillips and Kirby (2010):

> “The process in which the behavior of one individual is the product of observation of similar behavior in another individual who acquired the behavior in the same way”.

It is clear from this definition that iterated learning is only one special type of social learning mechanism, well distinct from teaching and imitation. It is not specific to language, but extends to any behavior acquired in this particular way, including music. It works for both peer-to-peer and parent-to-offspring transmission (Garrod, Fay, Rogers, Walker, & Swoboda, 2010; Theisen, Oberlander, & Kirby, 2010, Galantucci, 2009; Galantucci, 2005; Mesoudi, & Whiten, 2008). Finally, it is not unique to humans but also occurs in non-human animals (see Whiten & Mesoudi, 2008 for a review), most notably in songbirds (Fehér, Wang, Saar, Mitra, & Tchernichovski, 2009).

The effects of iterated learning on cultural evolution can be understood in the dynamics arising from cycles of observation, learning and use. One individual, or learner, induces (and stores in memory) the properties of a behavior that has being produced by another individual, the transmitter (*internalization*) (Griffith, Kalish, Lewandowsky, 2008; Kirby, Dowman, Griffith, 2007). The receiver then turns the role in sender and reproduces this behavior to a new receiver by reverse mapping (*externalization*). During the conversion from one domain to the other, apart from random noise, small, non-random changes are introduced in the original behavior due mostly to the limited processing capabilities of the animal brain, discussed earlier. When this process iterates, it exerts profound, predictable effects on the contents of the behavior transmitted. Small biases in information processing are amplified, and manifest in the structure of the cultural tradition transmitted (Kirby, Dowman, Griffith, 2007).

The iterated learning has been productively modelled in language evolution research with agent-based simulations (Brighton, Smith, & Kirby, 2005; Kirby, 2001; Kirby & Hurford, 2002; see also Kirby, Griffith, & Smith, 2007 for an exhaustive review), and only recently with laboratory experiments (Kirby, Cornish, & Smith, 2008; Tamariz & Kirby, 2015; Verhoef, Kirby, & de Boer, 2013; Kirby, Tamariz, Cornish, & Smith, 2015; Tamariz, 2017; Tamariz, & Kirby, 2016).

1.4.1. Iterated language learning: computational and experimental models

One aim in iterated learning research is to capture population-level changes in the system by modelling vertical transmission from peers to offspring. Computer simulations, supplemented by mathematical modelling, provided the first important insights on that matter. In simulations, population of agents endowed with particular learning algorithms, engage in a simple form of knowledge transmission: from one individual (the teacher) to another (the learner), along a chain in which each agent represents a ‘generation’. The learner in one iteration, becomes the teacher in the next. This type of transmission design is named *diffusion chain* (Bartlett, 1932; Esper, 1925). The *information bottleneck* (Kirby 2001) is also implemented in the model, so that the receiver is exposed only to a subset of data. The bottleneck wants to mimic an aspect of language acquisition that all children must face, the *poverty of the stimulus* (Chomsky, 1980). With the bottleneck in place, the learner is forced to generalize beyond their learning experience, to later reproduce expressions never encountered before. Results in the language domain support the hypotheses that cultural transmission lead to an increase of regularity and learnability. Initially unstructured languages, where sets of labels refer to structured meanings (colored moving objects), turn progressively to *compositional* ones (where the meaning of a complex expression depends on the meanings of its constituent parts and how these parts are arranged). The new peculiar arrangement is a key property of languages that makes them simultaneously highly compressible and expressive. Few strings can be systematically combined to produce a far larger amount of meanings. However, for that to happens, pressures must be exerted in the languages to counterbalance homonymy. Without this procedure, languages become systematic but ambiguous (see also next section). Similar findings were reported for *combinatorial* structure in phonological systems (de Boer, 2000), where few signal segments can be recombined to produce a larger amount of auditory signals. Compositionality and combinatoriality are two design features of human language (Hockett, 1960). The main contribution of computational modelling work was to show that these two major language features can develop without any intentional design by agents, but only as result of iterated transmission through the learning bottleneck.

Findings by computational models need however support. Learning and production algorithms of computational agents are based on very strong assumptions on the kind of mental processes at play in language acquisition and production. Therefore, the predictions generated by agent-based models need to be tested in the laboratory, i.e., replacing the agents with real human learners.

Inspired by the early work on diffusion chains (Bartlett, 1932; Kalish, Griffith, & Lewandowsky, 2007), and by more recent artificial language learning paradigms (Gómez & Gerken, 2000; Hudson, Kam, & Newport, 2005), Kirby et al. (2008) produced the first experimental version of iterated learning. In Kirby et al. (2008) the first participant in a diffusion chain learns an artificial miniature language system consisting of word-to-object mappings (*training phase*). He is then asked to generalize the mappings induced in the training phase to new objects (*testing phase*). Finally, a subset of the participant's output is used as training input for the next participant (or generation) in the diffusion chain (*transmission phase*), who repeats the same steps, and so on until the chain is complete. Critically, participants do not interact in the transmission phase, and remain unaware during the experiment that they are part of a diffusion chain. With such adjustments in place, any language change must result from unintentional adaptations by the language itself to repeated episodes of transmission. Over time, languages become more learnable and structured, as predicted by simulations (Kirby, 2001). Compositional languages emerged, where the meaning of a label was systematically predicted by the meaning of its subparts. However, this result was only obtained by removing homonymy from the participant's output, i.e., filtering out strings referred to the same label. This procedure was necessary to prevent the emergence of degenerate, ambiguous languages. Verhoef (2012) reported similar issues while modelling the emergence of combinatorial regularities in whistled sound systems (see for details next section and **chapter 2**). Overall, the iterated learning model (and diffusion chains methods in general) turned out to be a fundamental tool to reveal underlying cognitive and perceptual biases operating during cultural transmission (for iterated reproduction in single individuals see Jacoby & McDermott, 2017).

1.4.2. The evolution of musical structure in the lab: preliminary approaches

So far, I have described cultural transmission as a powerful mechanism to produce population-level linguistic phenomena. Given the formal similarities between language and music (Jackendoff, 2009), and the way they evolve and preserve over time, one may extend similar arguments to the musical domain (Merker et al., 2015). Music, like language, is a system culturally acquired and transmitted, based on a discrete set of particulate sounds, which are combined to form patterns with a great diversity of meanings (Merker, 2002). Like language, music exhibits a striking diversity of global and local structural variants (i.e., among and within cultures), with few of these variants being widely distributed among world cultures. The wide geographical distribution of these variants can be understood as cultural adaptation of music over historical periods to cognitive and perceptual constraints of the human brain. It is quite surprising then, that only few studies addressed this issue with empirical work.

1.4.2.1. Miranda, Kirby, and Todd (2003)

In a computer simulation, Miranda, Kirby, and Todd (2003) used the ILM to examine whether semantic compositionality can arise in a simple symbolic system, which consist of tunes expressing complex affective meanings. Although some disputable assumptions made by the authors, such as the language-like *propositional* character of the sound patterns, this work presents some good intuitions. Music, with other non-verbal expressive systems, is a communicative medium capable to affect our behavior (Sedlmeier, Weigelt, & Walther, 2011) and cognition, affective cognition above all (Cross & Woodruff, 2009). In the spectrum of musical affect we can find emotions and moods. If the mapping from sound onto emotions or moods (Koelsch, 2011; Patel, 2010) is arbitrary and open to cultural influences, one may ask to what extent compositionality can arise in symbolic systems where semantics is different from language (Arbib, 2013). The key finding of Miranda's work is that highly expressive compositional grammars emerge from initially unstructured ones as a result of repeated transmission through the data bottleneck. It is not clear to what extent the result was "pre-built" in the model: the learning algorithms were the same as in Kirby (2001, 2002) and the affective meanings, as the meanings in Kirby's work, already semantically combinatorial. However, it raised a fascinating question: can compositionality emerge in a semanticized melodic system, when this is culturally transmitted in the laboratory across generations of human learners? I address this question in the empirical **chapter 2**.

1.4.2.2. Verhoef (2012)

A study in the laboratory by Verhoef (2012) have also been judged as being informative for the cultural evolution of musical patterns, despite not being explicitly devised for this purpose. In this work participants use slide whistles to produce patterns of whistled sounds. The sounds generated by one participant are used to train the next in the diffusion chain, and so on. In the course of transmission, combinatorial regularities emerged, as predicted by computer simulations (de Boer, 2000). In the last generations, the set of signals was structurally rich, compressed and easier to learn and reproduce. Despite the emergence of music-like features (discreteness and contour), it is debatable whether this work is really about music. The instrument in use by the participants to produce signals was appositely devised to mimic an artificial phoneme inventory. It is difficult to disentangle the role played by constraints on perception and memory (Deutsch, 2012) and constraints on the production of signals. In Verhoef (2012), whistled sounds are relatively challenging to produce, because they require the modulation of sounds along a pitch continuum. Thus, mechanical biases for less costly actions may be entirely responsible for the emergence of musical features in this study (Tierney, Russo, & Patel, 2011). Last, mirroring Kirby et al. (2008), the results were only obtained by the use of an artificial procedure, preventing the participants from reproducing the same signal twice. Indeed, with only memory pressures in place, the whistle systems were expected to evolve to mono-melodic signal sets. This contrasts with the elaborate and diversified music we experience in everyday life, and suggests that other pressures (e.g., aesthetic, communicative, etc; Temperley, 2004; MacCallum, Mauch, Burt, and Leroi, 2012), in addition to learning, must be at work in the evolution of real music.

1.4.2.3. Ravignani, Delgado, and Kirby (2016)

A new experimental work on music evolution (Ravignani, Delgado, & Kirby, 2016) has recently come out, where the authors use the iterated learning paradigm to observe how rhythmic universals emerge over time. In this experiment, participants were provided with an electronic drum pad (connected to a computer) and a drumstick. They listened to 32 rhythmic (randomly generated) patterns via headphones (the *input*), and were then asked to reproduce them by memory using drumstick and pad (the *output*). Given the difficulty to memorize these patterns, some errors were introduced in the rhythmic system, which started to deviate from the originals. The erroneous output produced by the first generation of participants became the input for the next generation, and so on, so to form a transmission chain. After repeated episodes of transmission, the pattern grew increasingly structured, with in place regularities widely distributed across music traditions of the world, such as a tendency towards small integer interval ratios (especially 1:1 and 2:1) and a relatively steady beat (termed *isochronicity*) (Savage et al., 2015). This work represents the first attempt to grow music-universals in the lab (Fitch, 2017), and supports the view that different aspects of music structure can be understood as adaptations, through millennia of cultural transmission, to the human brain.

In summary, although cultural transmission research in the laboratory provides a suitable method for studying the evolution of patterns in music (as well as they do in language), there have been only few explicit attempts in this direction (Ravignani, Delgado, & Kirby, 2016; for an experiment with music signals and meanings see also Tamariz, Brown, & Murray, 2010). The studies just described do not provide a solid ground and need to be further validated with experimental research or other approaches.

1.5. From neural diversity to population structure

It should be now clear that one of the main current foci of experimental research on the cultural evolution of language is on the mechanisms and processes that generate central cultural invariants. A particular focus of these works is on the link that seems to exist between structural aspects of language and functional aspects of the human brain. Can we use similar arguments to explain the striking diversity of cultural variants observed in cultural symbolic systems? Humans are the example par excellence of cultural diversity in the animal kingdom. The diversity of symbolic systems such as languages and music in human cultures is quite astounding (Evans & Levinson, 2009; Nettl, 1983), with virtually any dimension free to vary, within the boundary of biological and physiological constraints (see Fitch, 2011 for an example with language).

The study of cultural variation has been, up to now, the proper domain of comparative methods, such as cultural phylogenetics. Phylogenetic methods were originally developed in biological science, specifically designed to detect and explain patterns and trends in biological evolution. Cultural evolution researchers (Dunn, Greenhill, Levinson, & Gray, 2011; Gray & Atkinson, 2003) use now these methods as rigorous, quantitative tools to reconstruct past patterns and trends in cultural evolution. The advantage of phylogenetic methods is the possibility to chart diversity in cultural traits, music traits included, across many societies, with relative cladistic evolution: the time the innovation arose, the location and the development in a comparative perspective (Rzeszutek et al., 2012). On the other hand, they only help to reconstruct cultural macroevolution patterns and trends. Given the several inherent limitations (e.g., time barrier, impossibility to manipulate variables of interest, etc.), these methods cannot really explain how innovation (and thus variation) originated, and originates, at first. While experimental methods, such as the ones described earlier, offer one means of addressing this problem (Mesoudi, 2007), they have only been focused on overall patterns of changes, neglecting specifically the individual *sources* of these changes (Caldwell, Cornish, & Kandler, 2016).

Following the considerations of the previous sections, we may believe that a potential source of innovation are the information processing capabilities of single individuals. Humans transform the information they receive from others, introducing features that fit with brain function and structure (Boyd & Richerson, 1985). Interindividual differences in brain function and structure as found in many studies (e.g., Zatorre, 2013), may manifest in difference in cultural behavior, at level of retention (transmission and innovation) and/or on the properties of the content (see the example of idiolects in language). According to this hypothesis, the pool of ‘neural diversity’ in a population would manifest in the form of frequencies of the cultural variants, which may spread in the population driven by different evolutionary forces, random or biased.

Random transformative forces are cultural ‘mutations’ (e.g., random copy errors, or sampling) and cultural ‘drift’ (random events that affect the distribution of cultural variants, especially in small populations). Other transformative, but directed forces are ‘biased transmission’ and ‘guided variation’. In biased transmission, some cultural variants are preferentially adopted rather than others, either because are easier to remember and reproduce (content-based biases), because the most frequent and available (frequency-based bias), or because produced by the most prestigious individuals (model-based bias). In contrast, driven variation does not affect the frequency distribution of the traits, but the traits themselves (Mesoudi, 2015). Cultural traits are modified when culturally transmitted, converging towards general constraints on learning and memory of the human brain (Boyd & Richerson, 1985; Sperber 1996). The peculiar characteristic of this evolutionary force is the nonrandom introduction of new variation.

Through these evolutionary mechanisms, cultural variants may spread in the population, possibly affecting its own structure. But how variation arises at first? Is it possible to relate cultural variant traits to the neural characteristics of the individuals? Which evidences do we have that neural variability manifests at population level? These points have been recently addressed by Dediu and Ladd (2007). The authors of this study found a correlation between the genetic population diversity of two brain growth genes, *Microcephalin* and *ASPM*, and the world distribution of tone languages. They proposed that these two genes determines individual differences in brain structure and function, and, in turn, small individual biases on the acquisition and processing of linguistic tones. These biases may be amplified during cultural transmission in populations with different genetic structure, and, finally, would manifest as population-level phenomena: as tonal (or non-tonal) languages.

Testing the hypothesis of a causal relation between individual and cultural diversity, is not a trivial task. It may be even harder to test with traditional cultural transmission methods, given the practical limits of an experimental set. First, it is necessary to change (unconventionally) the focus from population- to individual-level phenomena. I used the word ‘unconventional’ on purpose. Behavioral variability across individuals is considered in cultural transmission methods to be a source of noise, which must be removed by arithmetical procedures. Last, we need to provide an initial neurophysiological evidence that a link exists between cultural behavior and neural information processing (Dediu & Ladd, 2007). I address this issue in **chapters 3** and **4**.

1.6. Experimental research on cultural evolution: issues

So far I have described the iterated learning methodology as a fundamental tool to empirically address questions on cultural evolution. However, it also has important issues that deserve some attention. I summarize some of these issues in the following points.

The iterated learning model: First, despite the effectiveness in modelling the emergence of regularity, the ILM lacks of some of the fundamental aspects that characterize cultural transmission of symbolic systems, such as social learning, interaction and communication. In its classic version, design features can only emerge by the direct intervention of the experimenter (Kirby et al., 2008; Verhoef, 2012). This intervention is intended to mimic the adaptive pressures introduced by communication (Scott-Phillips & Kirby, 2010). Recently, Kirby, Tamariz, Cornish and Smith (2015) tested this assumption implementing communication and interaction in a novel version of the ILM. Two players are first exposed to the same artificial language (*training phase*). Afterwards they engage in a communication task, where roles of speaker and hearer are interchanged through repetitive rounds (*playing phase*). The output produced by these participants is later used as training input for the next generation of players (*transmission phase*). Pressures from both learning and communication led genuinely to compositional structure. Despite this, the present design does not allow to address directly the earlier assumption. Communicative pressures are here introduced *within* generation, not *between* (Kirby et al., 2008). Horizontal and vertical transmissions are found to be equally critical in language acquisition. Future versions of this model need to combine both factors if they want to provide more ecologically valid settings to test cultural evolutionary hypotheses on language (or other symbolic systems) (Tamariz & Kirby, 2016).

Neural constraints and population-level patterns: Second, the link between cognitive constraints and population-level patterns is based mostly on assumptions. One leading hypothesis in cultural transmission research is that regularization of symbolic systems results from adaptation to cognitive constraints of individuals' brains. However, the existence of these constraints has been only inferred by observation of behavioral data at population level (Kirby et al., 2008). No evidence has supported the existence of individuals' cognitive biases using neurophysiological investigation.

Empirical work in music evolution: Last, cultural transmission research, iterated learning in the specific, left the musical field largely unexplored. Barring few preliminary works (Miranda et al., 2003; Ravignani et al., 2016; Verhoef, 2012; Tamariz et al., 2010), this area lacks the organizational framework (concepts, approaches and methodologies) of historical linguistics.

The aim of my thesis is to fill these three gaps in cultural transmission literature: I introduce a model of cultural transmission of music, that involves alignment, communication and social learning between a sender and a receiver, the 'multi-generational signaling games' (or MGSGs) (Nowak & Baggio, 2016). I combine this model with neurophysiological methods to test whether some of the forms of cultural variation in auditory symbolic systems can be explained by individual information processing capabilities. Finally, I focus my investigation on two basic aspects of musical systems, namely melody and rhythm.

1.7. Research methods used in this thesis

This thesis includes research that employs both behavioral and electrophysiological methods. The behavioral method is a model of cultural transmission in the laboratory, the MGSGs (section 1.7.1). In the 'full' version of this model, I examine how musical structures evolve in an artificial (auditory) symbolic system as a result of transmission along diffusion chains of several generations (or participants). The use of a 'diffusion chain' design will create the condition to reveal the underlying cognitive biases at work in the cultural transmission of musical codes. In the 'reduced' version of the MGSGs, I restrict this process to single individuals: I examine how single individuals learn, transmit and restructure similar artificial symbolic system in two signaling games, the first played as receiver and the second as sender. The reduced version is used in combination with electroencephalography (EEG) (section 1.7.2). With a 'broken' diffusion chain design, I can capture the effects of cognitive biases on the music stimuli, and I may try to find a possible relation with the information processing capabilities of the individuals. The next section examines in greater detail these research methods.

1.7.1. Behavioral method

The model of cultural transmission and evolution in use here is a variant of the '*signaling games*' (Lewis, 1969; Moreno & Baggio, 2015; Skyrms, 2010), game-theoretic models typically employed in experimental philosophy and economics. In signaling games, a sender and a receiver coordinate to share information on a '*state*' of the world (e.g., "it is sunny outside"). 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 appropriate '*action*'. If the action does not match the observed state (e.g., umbrella), coordination fails. Crucially, in signaling games there is no mapping of signals to states that is given to or negotiated by the participants prior to the start of the game. The sender and receiver must therefore coordinate, through trial and error or other learning mechanisms, to develop a common code and successfully communicate. The extent to which each player adapts his own signal-state mappings to the other player reveals the division of coordination labor between the two players.

The field of study that investigates the emergence of novel communication systems via coordination and interaction is known as 'experimental semiotics' (Galantucci, 2009). In some of these works a 'closed-group' design is used (Mesoudi et al., 2008), where a constant group of participants (from simple dyads to microsocieties) repeatedly interacts in order to develop a communication system of signals referring to concepts (items). When the set of signals is open (i.e., the individual can produce any graphical sign), while states are closed and pre-established by the experimenter, these games are named 'semiotic referential'. They differ in organization from signaling games and the classic version of the iterated learning (Kirby, 2008), where states and signals are closed and pre-established by the experimenter ('semiotic matching games'). One of the most representative semiotic referential work is by Garrod, Fay, Lee, Oberlander, and MacLeod (2007), where a Pictionary task was adopted to study the emergence of conventions in artificial communication systems. There, dyads of participants (a 'director' and a 'matcher') were asked to develop a common code, where meanings were items (e.g., Microwave, television, etc.), and signals drawings depicting the items. The matcher (or receiver) was instructed to match different items of a list, with signals (drawings) produced by the director. With interaction, signals decreased in complexity (but only when a feedback was given to the matcher) becoming more arbitrary and symbolic, while communication accuracy increased. Repeated negotiation resulted in a progressive alignment between senders and receivers, with iconic signals providing the starting ground, and interaction promoting a shift towards less complex symbolic forms. To investigate the contribution of horizontal (dyadic interaction; Garrod et al., 2007) versus vertical transmission (Kirby et al., 2008), Garrod, Fay, Rogers, Walker, and Swoboda (2010) produced an experiment where graphical communication systems were vertically transmitted in chains in one condition, or negotiated by dyadic repeated interaction in the other condition: only horizontal transmission led to simplifications of signs. In addition, in that condition, matchers were in general more accurate in identifying the correct graphical signs. These results support the findings by Garrod et al. (2007). By contrast, the signals of the vertical transmission remained iconic. This result indicates that different pressures are at work during vertical cultural transmission (Kirby et al., 2008) and dyadic interaction: learning and expressive (or use) pressures respectively (Scott-Phillips, & Kirby, 2010). To what extent this result was given by differences in population size between the two conditions? Iterated learning experiments involve chains of multiple players, while in dyadic interaction the population is limited to two individuals. Fay, Garrod, Roberts, and Swoboda (2010) addressed this question, devising a community version of the game used by Garrod et al. (2007), a design known as 'microsocieties', to contrast to the simple dyadic condition. In microsocieties participants interact with the same player multiple times, then changing the partner and repeating the same game structure. In result of interactions and repeated partner's changes, the microsocieties developed signal sets of univoque representations (or 'global conventions'), shared between individuals of the same population, which were easier to acquire and reproduce. Different pairs in the 'dyad condition' developed instead idiosyncratic codes. This finding suggests that communities can evolve more efficient communication systems than isolated pairs.

The multi-generational signaling games (Nowak & Baggio, 2016) represent a variant of the models described. In line with the transmission chain design (Kirby et al., 2008) (section 1.4), the multi-generational signaling games require a replacement over time in the composition of the population. In this regard, multi-generational signaling games are useful to reveal and amplify small learning biases, otherwise hard to observe in single individuals. However, they also present the interaction and communication of lab-based models of experimental semiotics (Garrod et al., 2007; Garrod et al., 2010), which prove to be particularly useful for the emergence of stable cultural repertories. A similar design, which combines horizontal and vertical transmission (or 'chain of dyads'), was adopted by Kirby and colleagues (2015). There, horizontal and vertical transmission occur in two distinct phases. In this model, we integrate communication and transmission in a single phase. This design makes our model (very similar to) a replacement method (Tamariz, 2017). In multi-generational signaling games a two-player signaling game is played iteratively two times (reduced version) or until the diffusion chain is complete (full version). In the single signaling game, two players must converge on a mapping of signals (simple tone sequences) to meanings (simple and compound emotions), or '*signaling code*', via several iterated signaling rounds. Their roles are held constant throughout the session. Then, the receiver in one game becomes the sender in the next game. The new sender is asked to play a new game with the next participant (or '*generation'*; a new receiver), possibly retransmitting the code (i.e., the set of signal-state mappings) learned in the previous game. This design further distinguishes our method from the most recent versions of the iterated learning (Tamariz & Kirby, 2015; Theisen-White, Kirby, & Oberlander, 2011). There, intergenerational transmission is perfect by design, and changes are introduced only *within* games (interactive phase). In multi-generational signaling games, changes in the code can occur both *within* and *between* the games (i.e., in the sender's mind), and both can be captured by rigorous metrics. While the MGSGs fit the general description of the ILM (i.e. transmission to downstream generations), they also implement social learning, alignment and communication between users. Integrating horizontal and vertical transmission, multi-generational signaling games provide an interactional model of cultural learning, and a valuable alternative to the ILM.

1.7.1.1 A new laboratory model of music transmission

In my thesis, I use the multi-generational signaling games as a laboratory model of music transmission via coordination and alignment. Noticeably, this paradigm serves well the modelling of different aspects of orally transmitted music traditions, which is, still today, the primary mode of transmitting music in the world (Rubin, 1995; Kleeman, 1985). In a simple case of oral transmission, a musician (the sender) transmits through many repetitions his composition to another musician (the receiver) (Sawa, 2002). The accuracy of transmission may depend on the alignment of brain states between the two individuals, at least at two main levels of abstraction (Bharuca, Curtis, & Paroo, 2011; Temperley, 2004). The first level is structural. If the musical grammar used by the producer to generate music is detached by the listening grammar used by the receiver to represent that music in mind, the pair is not aligned, and the music will result difficult to understand, learn and appreciate (at least at a first hearing) (Lerdahl, 1992). Repeated interactions might be necessary for the producer to synchronize his mental grammars with the grammars used by the receiver, who, in turn, may decide to adopt these grammars to generate her/his own compositions (or to retransmit what had learned). Alignment in musical grammars may reveal a powerful mechanism for the intergenerational flow of musical structure. Linked to the first structural level, is the affective level. Music is not only a sonic pattern conveying information about its own construction, but it also carries an expressive affective content (Cross, 2005), a semantic association, prominently cultural (Meyer, 1956; Sloboda, 1985; Kivy, 1980; Tolbert 2001), with the structural content of the pattern. The composer or the performer might intentionally use this connection to align the listener (or multiple listeners) with his or her own brain affective states, or, instead, to drive them in a different direction. If the sender is successful, he can instill common emotional reactions in the listeners, an experience known as affective *resonance* (Bharuca et al., 2011). Social cohesion is one of the main reasons why people engage with music, and it is considered an important evolutionary function in the emergence of human musical behavior (Trehub, Becker, & Morley, 2015).

Another critical aspect of music oral transmission, is the asymmetry in information flow between senders and receivers (Sawa, 2002). Unlike verbal communication, music transmission is strongly unidirectional, with the roles of producer and listener maintained distinct during the performance (but not during music communication; see Dienes, Kuhn, Guo, & Jones, 2012). Information flow may be bidirectional to some extent - the positive or negative feedback given to a composer by the listeners - but the musical structure, and the semantic content that it bears, is always transferred from one generation of musicians to the next, and not vice-versa.

The third and most obvious aspect of oral transmission, is that music is passed across individuals without the help of notation (or any other external aid). Individuals must rely on their abilities to perceive and recall auditory patterns in order to maintain the material unchanged (Kleeman, 2012). In a long run, this mechanism of transmission may produce deep and predictable consequences on the evolution of musical structure (Trehub, 2015).

With the multi-generational signaling games, I model each of these three different aspects of oral transmission. Transmission of symbolic meanings (structural and affective) occurs via repeated interaction, alignment and coordination between players (Skyrms, 2010). The role asymmetries observed experimentally in signaling games (Moreno & Baggio, 2015; Nowak & Baggio, 2016) are expected to generate an information flow from senders to receivers, and from the first to the last generation of the transmission chains. Last, the transmission of symbolic codes depends in signaling games only on the ability of senders and receivers to process, store, and recall auditory symbolic patterns by memory.

1.7.2. Electrophysiological method

Electroencephalography (EEG) is a non-invasive method that measures the tiny electrical activity generated by neural processes. The high-temporal resolution of this method (in the order of milliseconds) makes it particularly suitable to investigate rapidly changing patterns of brain activity that underlie mental function, and more generally, brain information processing. Recordings can be made either of spontaneous activity (where they usually have a diagnostic value) or the activity elicited by some stimulation. In the latter case, a derivative procedure named ‘*event-related potential*’ (or ERP) (Luck, 2014) allows to isolate individual neurocognitive processes out of unspecific activity and electrical noise, by means of a simple averaging technique. The individual’s brain response to a single stimulus or event is not usually visible in the EEG recording, given the many sources of noise. The experimenter needs to present the same stimulus (or event) many times (hundreds or more), averaging the result together. In this way, the noise is averaged out, and the waveform that remains is known as ERP. The ERPs are visible as slow waveforms, and represent changes over time in voltage that are time-locked to specific events. Usually they reflect the weighted sum of several neural sources of activity, generated by the cognitive processing of the specific event. The most common way to quantify ERPs is by measuring mean amplitude and peak latency, which reflect respectively magnitude and timing of the examined neural process(s).

One important ERP component in auditory research is the '*mismatch negativity*' (or MMN) (Näätänen, Gaillard, & Mäntysalo, 1978). The auditory MMN is a negative, relatively automatic response elicited by an irregular event in a string of regular events. It typically peaks between 100 and 220 ms from the onset of the deviant stimulus. It reflects the brain's automatic detection of deviations from regularity representations, and its parameters ('*latency*' and '*amplitude*') seem to reveal how accurately these representations are encoded in memory (Alain, Cortese, & Picton, 1999; Winkler, 2007; Winkler, Denham, & Nelken, 2009). Previous studies reported a correlation between MMN parameters and behavioral accuracy (Kujala, Kallio, Tervaniemi, & Näätänen, 2001; Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993; Tervaniemi, Rytkönen, Schröger, Ilmoniemi, & Näätänen, 2001). This result shows that the MMN can be used as an index of (pre-attentive) auditory processing efficiency (Gottselig, Brandeis, Hofer-Tinguely, Borbély, & Achermann, 2004).

1.8. Thesis purpose and outline

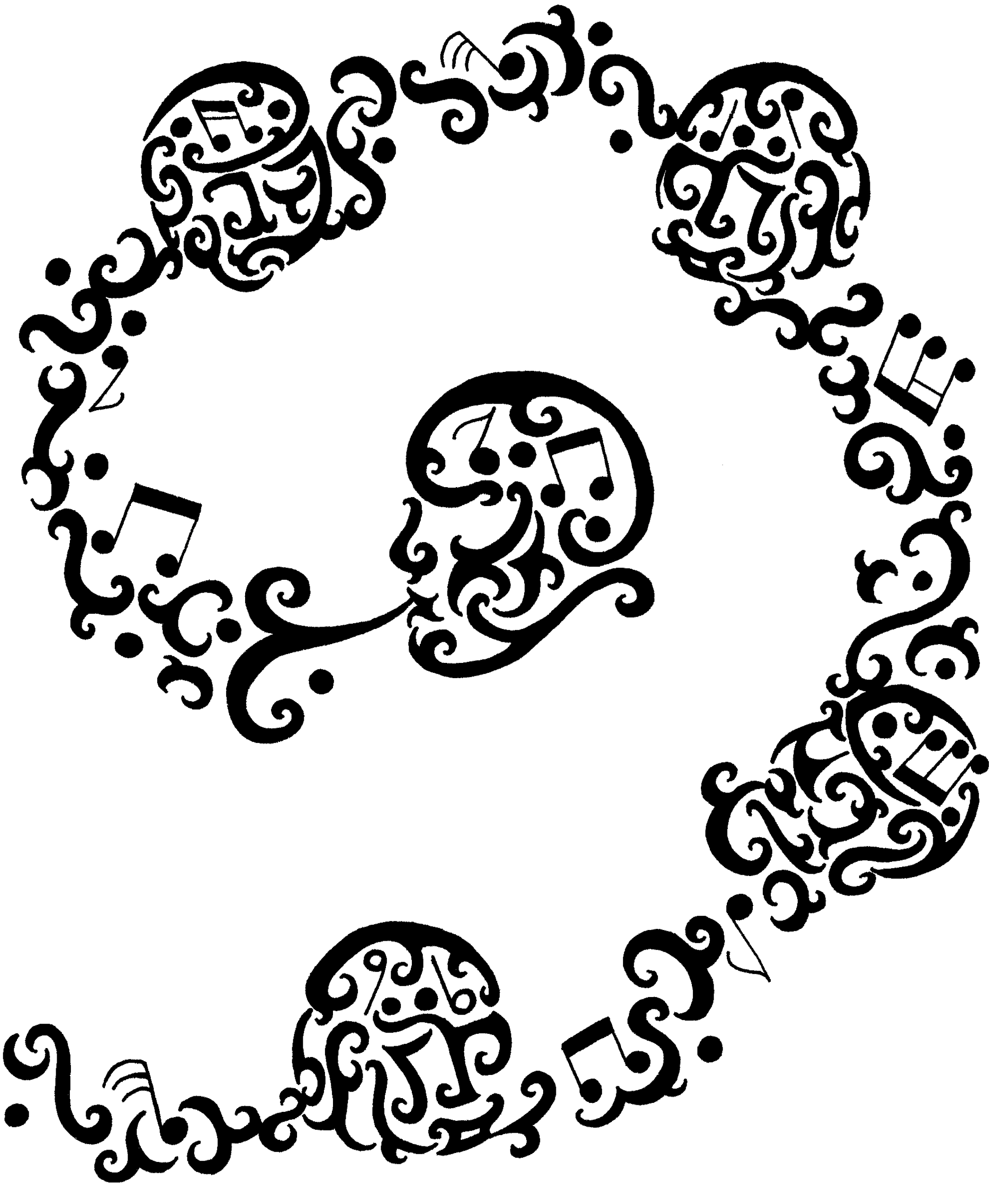
Our preliminary aim is to show that MGSGs are a viable laboratory model of cultural transmission via alignment and communication for auditory symbolic codes. I expect this model to be effective in reproducing two typical traits of cultural evolution (Tomasello, 2009): directed changes of cultural traits over generations, and asymmetry in information flow from senders (or transmitters) to receivers (or learners).

The main aim of this thesis is to provide initial, experimental support to the idea that adaptation to cognitive constraints of the human brain can explain, at the same time, some of the forms of *invariant* and *variant* aspects observed in auditory symbolic systems of different cultures, music in particular. My first aim is to support the view, with laboratory work, that some of the near-universal properties of music can be directly explained as interplay between cognitive and social processes. Second, I aim to provide an initial evidence of the relation between cultural adaptation and neural information processing. I do this by focusing on two major structural components of music, melody and rhythm, using a combination of behavioral and neurophysiological methods. These two subcomponents will be studied separately, as it is done in much of the experimental literature on music psychology (Deutsch, 2012). Critically, I expect that no artificial procedures (Kirby et al., 2008; Verhoef, 2012) will be here necessary for the emergence of musical properties. The aims of this thesis will be explored in 4 experimental works, two behavioral and two EEG studies, described in 3 empirical chapters (chapters 2-4).

The first two empirical chapters focus on the melodic dimension of music. **Chapter 2** examines *how* artificial tone systems evolve as a result of being repeatedly learned and transmitted in populations of human learners. Specifically, I explore how constraints on melodic perception and memory shape a miniature artificial system of isochronal tone sequences endowed with affective meanings during transmission. To this purpose I implement the MGSGs, in which each transmission step introduces a new generation of senders and receivers. The aim of chapter 2 is to gain insights on the mechanisms and pressures at work in the evolution of melodic structures.

To follow up on this, **chapter 3** explores whether inter-individual differences in neural information processing may account for variation observed in three key processes underlying cultural evolution: social learning, transmission and structural modification. To this purpose I combine behavioral and EEG methods in two successive days. Specifically, I test whether individual differences in auditory information processing, as revealed by event-related potentials (ERPs) recorded on day 1, predict acquisition, transmission and modification of melodic signaling systems on day 2. This chapter provides the first neurophysiological evidence of cognitive constraints at work on cultural transmission.

In **chapter 4** I replicate and extend the results of chapters 2 and 3 to the temporal dimension of music, namely to rhythm. First, I address whether the emergence of one specific feature of musical rhythms, beat regularity (or *isochronicity*), can similarly be explained in terms of repeated transmission constraints. To this purpose, artificial miniature rhythmic systems are culturally transmitted in the laboratory through MGSGs (see **chapter 2**). Then, I examined whether temporal regularization has neural origins, and if differences in neural temporal processing across individuals relate to differences in rhythmic behavior and structure. To this purpose, I tested whether learning, transmission and temporal regularization of similar rhythmic systems can be predicted by using individual neurophysiological data (see **chapter 3**). Finally, a summary of the main findings with a broader discussion is presented in **chapter 5**.



Chapter 2

Cultural evolution in the laboratory of melodic structures

Lumaca, M., Baggio, G. (under review). Cultural transmission and evolution of melodic structures in multi-generational signaling games.

Abstract

It has been proposed that languages may evolve by adapting to the perceptual and cognitive constraints of the human brain, and by developing, in the course of cultural transmission, structural regularities that maximize or optimize learnability and ease of processing. To what extent do perceptual and cognitive constraints similarly affect the evolution of musical systems? We conducted an experiment on the evolution of artificial melodic tone sequences using multi-generational signaling games as a laboratory model of cultural transmission. Signaling systems, using 5-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 signaling systems, following principles of proximity, symmetry and good continuation. Whereas 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 universal cognitive and perceptual mechanisms operative in the cultural evolution of musical systems, which may however differ from the mechanisms at play in language evolution and change.

2.1. 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 (Christiansen & Chater, 2008; Deacon, 1997). During cultural transmission, languages change as the result of pressures to adapt to the constraints imposed, among others, 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 re-transmit (Kirby et al., 2008; Kirby, Griffiths, & Smith, 2014; Kirby & Hurford, 2002). One may ask whether and to what extent similar principles extend to the evolution of melodic systems, such as music and vocalization (Patel, 2010). Universal acoustic and auditory mechanisms may explain the wide geographical distribution observed for a narrow set of music features, or music statistical universals (Savage et al., 2015). Although the idea is not new (Merker, 2006; Merker et al., 2015; Terhardt, 1987; Trehub, 2000), only a few studies have addressed it by means of laboratory experimentation (Verhoef, 2012).

2.1.1. Iterated learning and beyond

In a study by Verhoef (2012), the emergence of phonemic 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 (Kirby et al., 2008). In the course of transmission, combinatorial regularities emerged. In the last generations, the set of signals was structurally rich, compressed and easier to learn and reproduce.

Despite the effectiveness of iterated learning in modelling the emergence of structural regularities in artificial miniature languages (see Kirby et al., 2014 for an overview), the model in use by Verhoef (2012) requires the direct intervention of the experimenter in order to filter out redundant signals from training sets: mono-melodic signal sets were otherwise expected to emerge. These interventions were intended to mimic the adaptive pressure imposed by communication (Scott-Phillips & Kirby, 2010) which is by design absent in this version of iterated learning (but see Kirby et al., 2015 for a new procedure; see Nowak & Baggio, 2016 for a critical discussion).

Here, we investigate the evolution in the laboratory of a miniature artificial tone system, where auditory signals refer to distinct meanings. We assess whether cognitive and perceptual constraints shape auditory codes so that they gradually acquire features that would enhance acquisition and memory retention (Mesoudi & Whiten, 2008). To this end, we test whether artificial tone systems would evolve toward more compressible and simpler forms (Tamariz & Kirby, 2015) using measures of compositionality (Kirby et al., 2008) and melodic compression (see Cornish, Smith, & Kirby, 2013 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 (Deutsch, 2012).

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 (or scale) (Tamariz, Brown, and Murray, 2010). This is cognitively relevant for both music (Deutsch, 2012; Patel, 2010) and speech intonation perception (Mertens, 2004); in addition it allows us to introduce quantitative, versus qualitative (Verhoef, 2012), analyses to measure the evolution of melodic structures. Second we present and apply a model of cultural evolution (signaling games) in which the transmission and acquisition of the codes occurs via repeated rounds of interaction between senders and receivers. While this is not a novel approach in language evolution research (Caldwell & Smith, 2012) (see especially Nowak & Baggio, 2016), only one study has exploited it in the auditory domain (Lumaca & Baggio, 2016). We expect structural regularities to arise spontaneously as a result of coordination and communication, possibly reflecting learning 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 (Patel, 2010). 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 may be applied to musical melody (Fitch, 2005). There is agreement among musicologists and cognitive scientists that musical phrases (and smaller or larger units of musical discourse) can be used, and have historically being used, to convey a variety of types of meaning, including affective meaning (Davies, 1994; Koelsch, 2011; Koopman & Davies, 2001; Leman, 1992). If music fits the properties of a communication medium (Shannon & Weaver, 1949; Cross, 2004), with a sender (the musician) transmitting some message to a receiver (another musician) through some channel (air), we can adopt the terms ‘musical code’ and ‘signal’ for the objects of transmission. In this context, we will refer to signals as a set of sonic patterns, and musical code (or code) as the set of mappings between signals and meanings in the mind of the two players. The goal of the players is to align their codes, possibly to the one used by the sender, so to have a successful transmission of information (musical properties referring to meanings) from one player to another (section 2.1.2). In this respect, 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 Hsieh, Hornberger, Piguet, & Hodges, 2012;

Palmer, Schloss, Xu, & Prado-León, 2013; and especially Gordon, Key, & Dykens, 2014). This will allow us, following Kirby et al. (2008), to explore whether compositionality, whereby the meaning of a signal is composed by the meaning of its subparts, arises in the organization of symbolic pitch patterns too.

2.1.2. The present study

In signaling games (Lewis, 1969; Skyrms, 2010), 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 appropriate action. If the action does not match the observed state, coordination fails. Crucially, in signaling games there is no mapping of signals to states which is given to or negotiated by the participants prior to the start of a game. The sender and receiver must therefore 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 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 (Schönpflug, 2008), which may also be relevant for the evolution of auditory symbolic systems (Temperley, 2004). In classic versions of iterated learning (Kirby et al., 2008; Verhoef, 2012), no coordination process may occur between players, and the 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 (Scott-Phillips, Kirby, & Ritchie, 2009; Selten & Warglien, 2009). 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 & Baggio (2015) 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 which tends to become the common code. While signaling games with fixed roles show a limited external validity as communication models, they are a viable model of cultural transmission (Nowak & Baggio, 2016), serving particularly well the transmission of music, where the information flow between a sender (music producer or performer) to a receiver (another musician) is strongly asymmetrical i.e., from the first to the second.

In our experiment, the states (possible messages) to be denoted by signals are emotions, shown on the computer screen as photographs of one actor's facial expressions (male or female), and signals are sequences of 5 pure tones drawn from the Bohlen-Pierce (BP) scale (Mathews, Pierce, Reeves, & Roberts, 1988). The group of signals belonging to the same code, and denoting different emotions, will be referred from now on as ‘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’ task is to develop and finally agree on a mapping of signals to states over successive signaling rounds (Fig. 1). The initial code (or seed) is produced by the experimenters. Seeding is a common technique in experimental studies of cultural transmission (Flynn & Whiten, 2008; Whiten et al., 2007), which allows the experimenter to seed controlled properties on the initial material, then observing how they spread and modify as a result of transmission. The properties controlled in our seed were the melodic direction and the distance between adjacent tones (respectively, ‘melodic contour’ and ‘proximity’), and the systematicity in mapping between signals and meanings (or *compositionality*) (section 2.2.6). 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 participant (Generation 2). These two participants play respectively as the sender and receiver, with their roles held constant throughout a session (Fig. 1). 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). Multi-generational signaling games are thus played sequentially and are organized as in a vertical diffusion chain consisting of 8 generations.

As a preliminary result, we expect to further validate signaling games as a model of the cultural transmission of symbolic affective systems (Lumaca & Baggio, 2016; Moreno & Baggio, 2015; Nowak & Baggio, 2016). We assessed the direction of information flow in the chains using two indices: role asymmetry and coordination (Moreno & Baggio, 2015). When players share most mappings at the end of a game, such 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 flown forward in the chain, from senders to receivers.

In addition, we test how constraints on compression of information (Chater & Vitányi, 2003) and on melodic perception and memory (Deutsch, 2012; Snyder, 2008) shape the structure of tone sequences during transmission. First, we apply measures of compositionality and melodic compression. These measures can assess how similar signals of the same set are (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 (Kirby et al., 2008; Tamariz & Kirby, 2015). Learning can also be affected by peculiar structural features embodied in melodic segments, or grouping features, which are known to facilitate the encoding in memory (Deutsch 2012). 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.

2.2. Methods

2.2.1. 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 2 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.

2.2.2. Diffusion chains

Signaling games were organized in 8 vertical diffusion chains of 8 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.

2.2.3. States

The states or items, denoted by signals in signaling games, were 5 emotion categories of different complexity: 3 basic and 2 compound emotions, represented as facial expressions (Fig. 2). Although the perception and production of compound facial expressions is consistent with the subordinate basic categories they are built upon (Du, Tao, & Martinez, 2014), they are perceived as independent expressions. Following the coding system for facial expressions developed by Ekman and Friesen (2003), facial expressions of basic emotions (peace, joy and sadness) were analyzed in 2 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, we determined a meaning space of 2 dimensions: upper face and lower face. The upper face was coded using two variants (open or closed eyes), and the lower face involved three variants (mouth corners up, straight and down).

2.2.4. Signals

The constituent 5 tones of the signals were drawn from the Bohlen-Pierce (BP) scale (Mathews et al., 1988). The number of tones fits the size of melodic groupings (Snyder, 2000), the basic units of melodic perception. In the equal-tempered version, a tritave (3:1 frequency ratio) is logarithmically divided into 13 equal steps (Suppl. Fig. 1), larger in size than the corresponding Western semitones (146 cents vs 100 cents). This makes BPS a ‘macrotonal’ scale. ‘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: $F = k * 3 (n/13)$; 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 as to maximize the number of low integer frequency ratios between any combination of tones (Loui, Wessel, & Kam, 2006). Sounds were sine waves 500-ms long, with 50-ms fade in-out and 50-ms of inter-tone interval (see Suppl. Fig. 2). 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 section 2.2.7 (Compositionality).

2.2.5. Procedure

Prior to the start of a game, each player in a pair was independently trained to associate each facial expression to one simple or compound emotion, in 3 blocks. In the first and second blocks, respectively, facial expressions for simple and compound emotions were randomly presented on the center of the screen to the participant, one at time. Given each facial expression, the participant had to choose the appropriate emotion (e.g., peace+joy) among a vertical list of orthographical options (i.e., peace, joy, sadness, peace+joy, peace+sadness), 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 test in block 3, the entire set of facial expression was presented again, one at a time, 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 in 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. Fig. 1 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 sec duration); a blank screen followed. Next, the sender was asked to generate a 5-tone sequence from the pool of 5 tones available using the keyboard, to signal to the receiver the state he had seen. All tone sequences were isochronous: only 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 5 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 5 facial expressions shown on the screen, i.e., the one he thought the sender had seen, by using the '1'-'5' keys 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 (i.e., 50 positive feedbacks), with one break at 25 correct trials.

2.2.6. Seeding

We designed 4 sets of 5-tone sequences (Suppl. Fig. 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 5 blocks of increasing complexity and duration. In each block, he was instructed to produce a 5-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.

2.2.7. 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 (Siegel, 1956) 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 (Winter & Wieling, 2016) in R Studio (R Studio Team, 2015) with the package ‘lme4’ (Bates, Maechler, & Bolker, 2012). The same approach was used in previous research on iterated learning (e.g., Verhoef, 2012). 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 modelled 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. Indices 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 $\alpha = 0.05/5 = 0.01$.

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 Hamming distance (Hamming, 1986), here the number of string elements shared between 2 strings of equal length. In our analyses, the strings were either 5-tone sequences (tone distance) or their contour transforms, i.e., 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 (Dowling, 1971).

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, or the trial from which sender and receiver use the same mapping consistently until the end of the game (barring random errors) (Moreno & Baggio, 2015). More specifically, we first identified a block of 5 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 indices of model efficiency were used: 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). Finally, 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).

Transmission and Innovation

The tone and contour similarity measures (modified Hamming distance) between sequences denoting the same emotion, produced by players of adjacent generations, are indices of the faithfulness of transmission (a between-player measure). A code that is faithfully transmitted have values close to 1; vice versa, values close to 0. 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 indices do not mirror each other. Values close to 1 suggest that the code were restructured between games. Four values (2 for transmission, 2 for innovation) were measured for each signal set in the experiment.

Proximity

A reliable measure of pitch proximity is the absolute mean interval size (Von Hippel 2000), 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 tones sequences of the same set. Low values are associated to sets with more proximal melodies, and high values with more distant melodies. The actual statistical distribution of unison, small (1 or 2 macrotone 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 provides a further, different measure of melodic size.

Measures of mirror forms

We tested the evolution of signals towards three types of transformations: retrogrades, inversions (mirror forms) and retrograde inversions (Dowling & Angeles, 1972). 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, such 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, 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.

Measures of contour smoothness

The Shannon entropy (Shannon, 1949) of contour transforms was obtained with the following equation:

$$H(X)= -\Sigma P(xi)\log 2P(xi)$$

where X is the string of constants, ups and downs and P(xi) is the probability the 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.

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 on this measure see Tamariz & Smith, 2008). Following Cornish et al. (2009), signals from the last pair of each chain were partitioned into segments, such that elements could be reliably associated to meaning dimensions. As result of this, signals were segmented into units of 2 and 3 tones. We applied this segmentation to the entire pool of signals produced in the experiment. Then, we followed Tamariz (2013) 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.

Melodic compression

Melodic compression measures the similarity between signals of the same set (Cornish et al., 2013). 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 (mono-melodic or mono-contour signal sets).

2.3. 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.

2.3.1. 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 (Lumaca & Baggio, 2016; Moreno & Baggio, 2015; Nowak & Baggio, 2016).

2.3.2. Transmission and Innovation

Fig. 3 shows the changes in the rates of transmission and innovation as measured by tone (Fig. 3a) and contour distances (Fig. 3b). Using the latter measure, we found an increase in transmission between the second and third generations (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 generations (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=±0.02), while innovations introduced by participants remained frequent in successive generations (mean contour distance=0.19, SD=±0.02).

The observed increase in transmission fidelity (Kirby et al., 2008; Kirby et al., 2015) between the second and the third generations suggests that some critical changes in code structure occurred in the output of generation 2. From this point on, gradual changes occurred, as indicated by lower innovation values. The specific direction of such changes is described below (see Suppl. Fig. 4 for an example of evolving code).

2.3.3. Proximity

An initial, non-significant 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) (Fig. 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).

2.3.4. Interval distribution

To test whether changes in proximity are due to a progressive increase in the frequency of monotone melodies, we performed a two-tone transition analysis. Comparisons of the first and last generations, 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 (Suppl. Fig. 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).

2.3.5. Melodic transformations

A systematic rearrangement of regular patterns occurred through vectorial operations performed by players 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, $p<0.001$) (Fig. 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) (Suppl. Fig. 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.

2.3.6. Contour smoothness

We also tested the emergence of smoothed melodic lines 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. Fig. 4b shows changes in this measure across generations. The decrease between generation 1 and 2 was not significant (n=40, Z=-1.723, p=0.08), and the same applies to generation 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 change here ($\chi^2(1)=2.87$, p=0.09). This indicates that the initial melodic material, that mostly consisted of elaborate contours, was replaced by more regular, predictable patterns (Suppl. Fig. 6), possibly with different factors at work in producing the observed result (see Baseline).

2.3.7. Melodic compression

Fig. 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 generations 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 generations (G1/G7, n=8, Z=-2.533, p=0.01; G2/G7, n=8, Z=-1.065, p=0.28).

2.3.8. 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 re-use of fewer smaller subsequences to produce whole sequences. Possibly, the combinatorial recycling of these subunits may be not fully random but also related to the structure of the meanings conveyed. Our subunits were melodic segments of two and three tones, while upper and lower face were the meaning dimensions possibly related. This feature is known as compositionality, and its evolution is shown in Fig. 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).

2.3.9. Introducing baselines

To test whether the observed structural regularities evolved by chance 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 random-like state (p>0.05 for all comparisons between original generation 1 data and shuffled data), proximity (n=8, p=0.02; baseline B=3.46), mirror forms (n=8, p<0.01; B=0.35), 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 (Figs. 4a, 4c and 4d). 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.

2.4. Discussion

Our results suggest that an artificial system of tone sequences endowed with semantics tends to regularize in several non-trivial ways when it is transmitted across generations. We observed the emergence of several structural features that may promote learning and memory retrieval. In particular, proximal melodies with symmetric patterns of tones emerged over time. In addition, the similarity between melodic segments increased from the first to the last generations, whereas changes in compositionality were negligible and remained around baseline levels. In contrast with earlier iterated learning work on melodic transmission (Verhoef, 2012), no external intervention on the participant output was necessary to produce these results.

2.4.1. Signaling games as a model of cultural transmission

The initial sharp increase in transmission observed in our data set contrasts with the slow changes of previous iterated learning studies (Kirby et al., 2008; Verhoef, 2012). 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 receiver's learning (Tan & Fay, 2011), leading to a rapid convergence to a shared symbol-state mapping (Swarup & Gasser, 2009). As a result, a more faithful reproduction of the code is expected after 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 (Caldwell & Millen, 2008). In agreement with Moreno & Baggio (2015), 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 thus from the first to the last generation of the diffusion chains.

Cumulative transmission and asymmetry in information flow, as found in our experiment, are essential traits of cultural evolution (Tomasello, 2009). These properties make signaling games a viable laboratory model of the cultural evolution of auditory tone systems (Tamariz, Brown, and Murray, 2010), that combines communication and transmission of iterated learning (Kirby et al., 2015), with a replacement method (Caldwell & Smith, 2012). Multi-generational signaling games can be used as a complementary approach to iterated learning (Kirby et al., 2015).

2.4.2. 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 (Deutsch, 2012). 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 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 (Dowling & Angeles, 1972; Schoenberg & Newlin, 1951). By using high-order abstractions of a given melodic sequences, such as transforms, participants introduced moderate levels of redundancy in a signal, 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 grouping (e.g., Rohrmeier & Cross, 2013; Rohrmeier, Rebuschat, & Cross, 2011).

These results suggest that melodic information is recalled as particular instances of grouping or configuration, but they also provide indications on how these configurations may be encoded and recalled. We noticed that fine-grained information was usually lost, whereas the contour of melodies was preferentially encoded (Dowling, 1994). As a first result, we found that code transmission was enhanced, as revealed by melodic contour changes rather than pitch sequence Second, the signal set was reorganized in fewer melodic shapes, with the resulting increase of similarity between signals of the same set. A analogue finding was reported by (Bartlett, 1932) in serial production experiments. In that study he reproduced at a small scale the oral tradition of a folk tale. He discovered that only ‘dominant’ features were well remembered in retellings, while details were omitted or modified. Based on this result, Bartlett proposed that remembering is more a reconstructive process based on a high-level abstract model than an exact replication of the 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 multi-level chunking mechanisms to avoid it being overridden by incoming information. After this process, new memory represents just an abstract summary of the original sensory input (Christiansen & Chater, 2016). Results in our experiment show that similar chunk-and-pass mechanisms operate in the evolution of melodic material, probably affecting the transmission and the evolution of melodic structures, historically documented in the musical domain (Kleeman, 1985; Sawa, 2002). The qualitative observations by (Kirby et al., 2008; Verhoef, 2012) 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 speech domain. The arrangement of sounds in intervals of smaller size (proximity), with coherent directions, is a well-documented phenomenon in different musical (Huron, 1996) and linguistic cultures, when continuous pitch glides are converted into discrete pitch patterns (Huron, 1996; Patel, 2010). The “[...] greater relative emphasis on global features […] than on local details” (Trehub, 2000), and likewise the proportion of transformations found in our experiment can be also commonly found in different cultures and historical periods (Dowling, 1971; Dowling & Angeles, 1972; Huron, 1996).

Studies using iterated learning show that cultural systems, when repeatedly learned and re-used, are regularized so that structure becomes compressible, or easier to acquire or reproduce. Compression is a fundamental cognitive principle (Rissanen, 1978). 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. When this principle is applied to languages, and when also expressivity comes into play, it is partly reflected in a key feature of (islands 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 (Kirby et al., 2008; Verhoef, 2012). The negligible changes in the degree of compositionality found in this experiment stand in contrast with previous iterated learning studies (Kirby et al., 2008; Theisen et al., 2010; Kirby et al., 2015), and are instead consistent with an arrangement of patterns of sounds based on perceptual principles (Fitch, 2006; Pierrehumbert & Hirschberg, 1990). In support of this conclusion, a relatively inefficient transmission of compositional structures was reported in musically naive participants for a miniature artificial tone language (Tamariz, Brown, & Murray, 2010). We cannot exclude that the present outcome might change if a full compositional semantics, as in Kirby et al. (2008), is 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 code 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 mono-melodic systems. The tendency of signaling systems towards homonymy has been reported in previous iterated learning studies (Verhoef, 2012; Kirby et al., 2008). 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 (Chater & Vitányi, 2003; Tamariz & Kirby, 2015), 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 progressively limited set of melodic elements (Adams, 1976). 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 re-used 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 re-use of these elements, or groups of elements (motives; Bartel, 2006), to produce a virtually boundless number of complex musical phrases (Verhoef, 2012). 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 are not necessarily mutually exclusive.

2.5. Conclusion

Our results suggest that culturally transmitted artificial tone sequences are subject to the same basic set of cognitive and perceptual constraints that operate on the cultural evolution of auditory symbolic systems, musical and speech melodies included. 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, music in particular. The present work is the first experimental contribution to the debate on the role of cognitive constraints in the cultural evolution of tone patterns. It provides experimental evidence that basic principles of auditory organization as well as cognitive constraints can drive the evolution of simple tone sequences toward universal melodic regularities.

2.6. Figures and tables

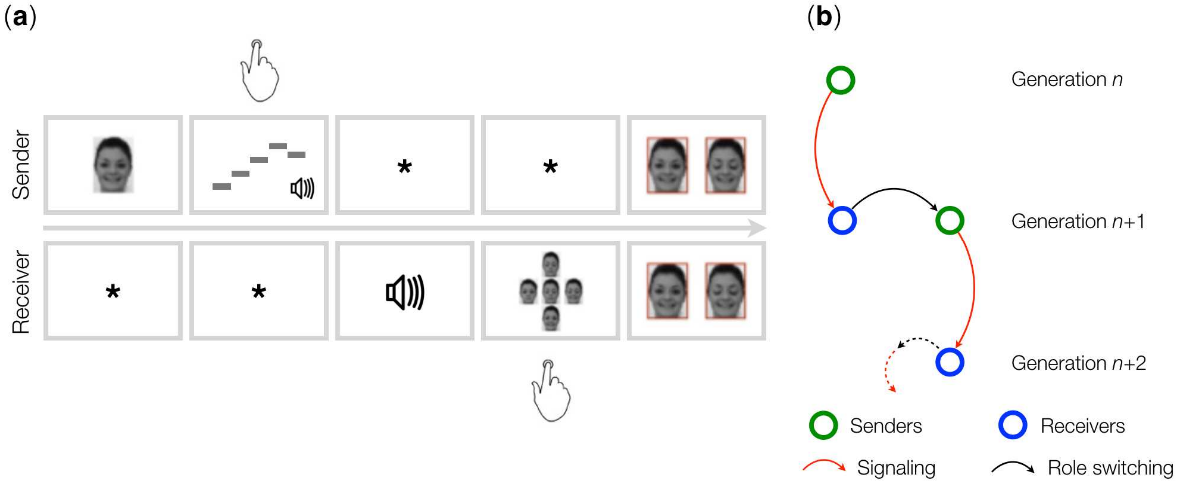


Fig. 1. a) A trial of a signaling game with 5 emotions (3 simple, 2 compound) as states (or messages) and 5-tone sequences as signals. Emotions are shown as expressive faces. Tones are drawn from the Bohlen-Pierce scale. The top and the bottom rows show what the sender and receiver view on their screen, 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 framework, while a red frame is used for negative feedback. b) Structure of multi-generational signaling games. The receiver in one game (generation *n*) becomes the sender in the next game (generation *n*+1), and diffusion chains are thus constructed.

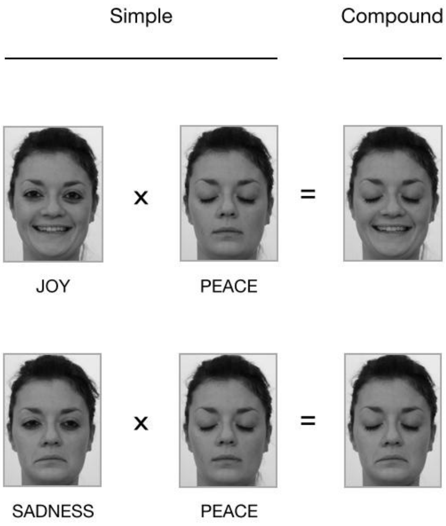


Fig. 2. Stimulus set of the study on emotion expression. States are faces expressing either simple (sadness, joy, peace) or compound emotions (peace×joy, peace×sadness). Stimuli for compound emotions are constructed by merging the upper portion of ‘peace’ with the lower portion of ‘joy’ or ‘sadness’.

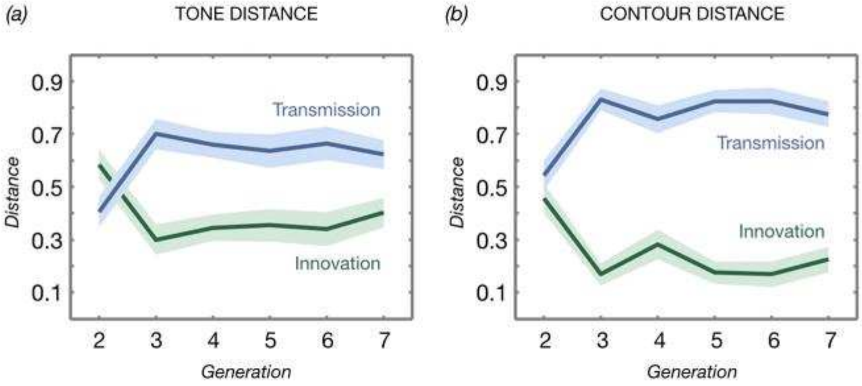


Fig. 3. Changes in transmission and innovation over generations as measured by (a) tone and (b) contour distance. The dark line shows the mean value for each generation across chains; the blue-shaded area represents SEMs. Transmission is the modified Hamming distance between codes produced by senders of adjacent generations (between-players measure). Innovation is calculated as 1-D, where D is the distance between the code learned and the code produced in the following game by each participant (within-player measure).

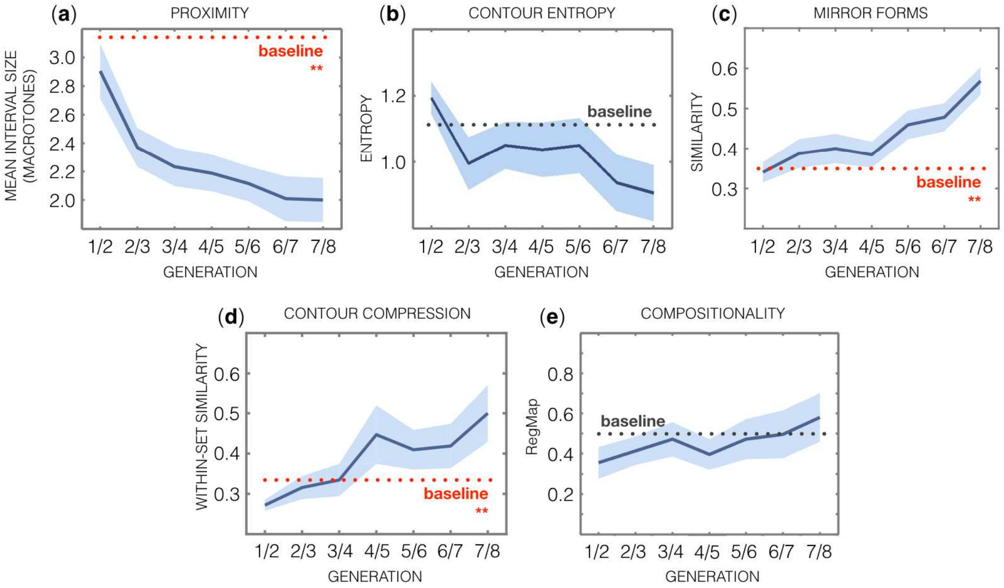
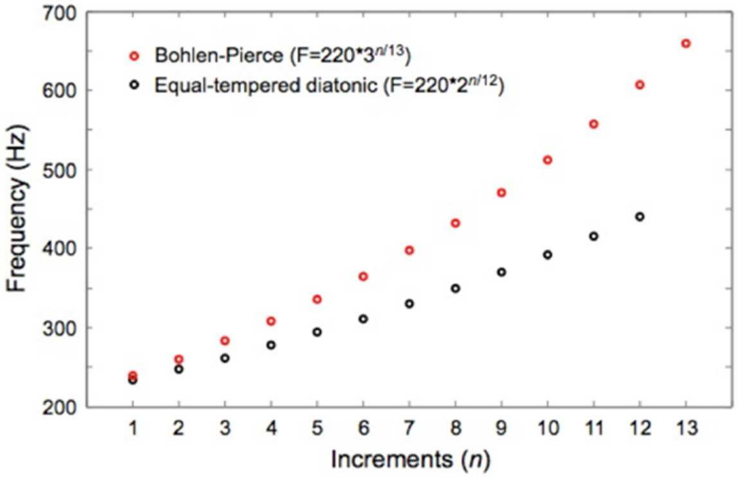
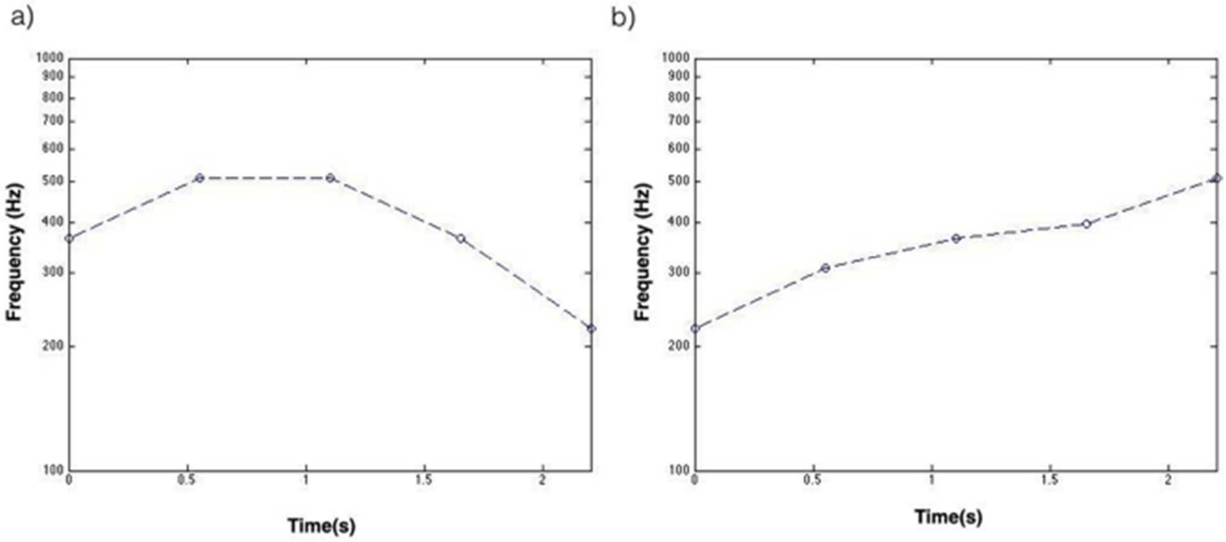


Fig. 4. Temporal evolution of structural regularities (a-e) across generations. The dark line is the average value computed for each generation across chains; the blue-shaded area represents the SEM.

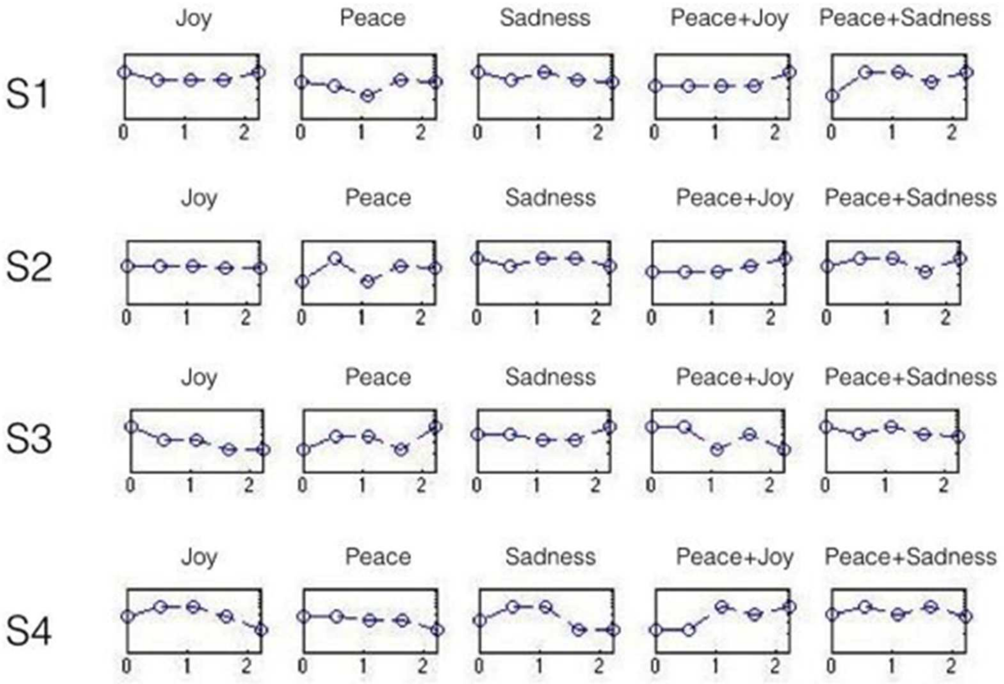
2.7. Supplementary material



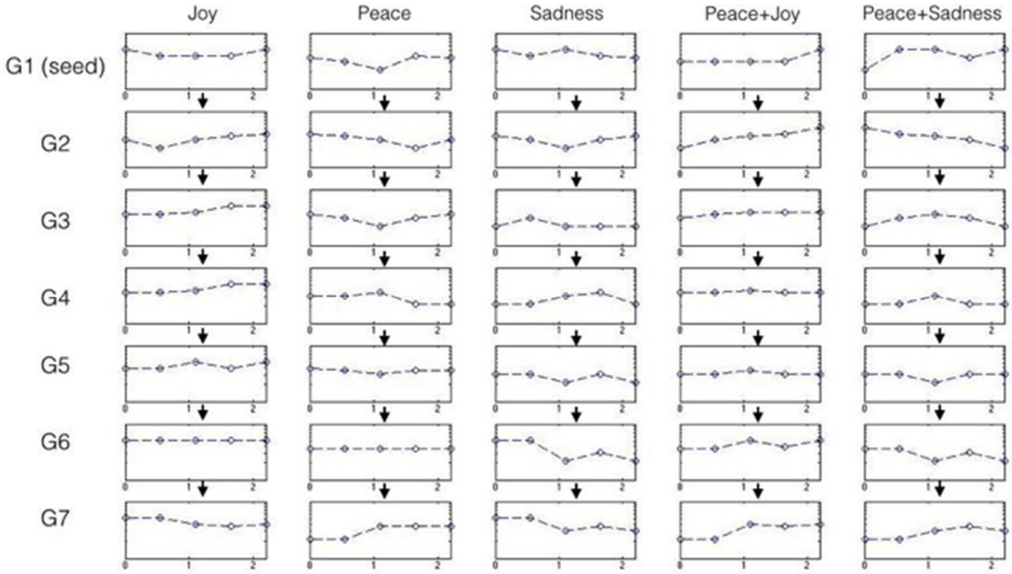
Supplementary Figure 1. Frequency increments in the Bohlen-Pierce system (red) and in the equal-tempered diatonic system (black).



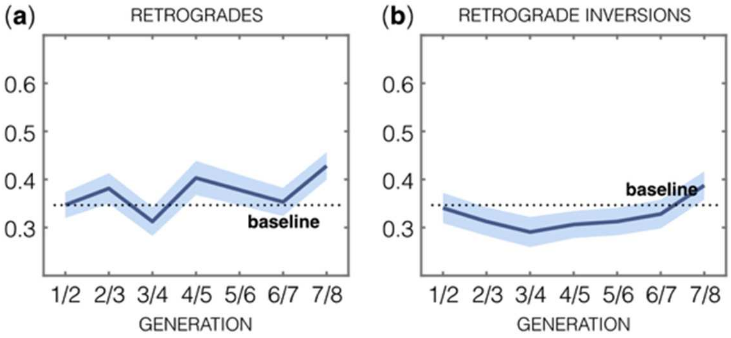
Supplementary Figure 2. Diagram-notations of two example tone sequences. Each sound is 500-ms long with 50-ms of fade in-out and 50-ms of inter-tone intervals. Time flows from left to right. The y-axis is in log-scale. Circles represent the onset of pitches, while the dotted line is used to represent the contour of the melodic pattern.



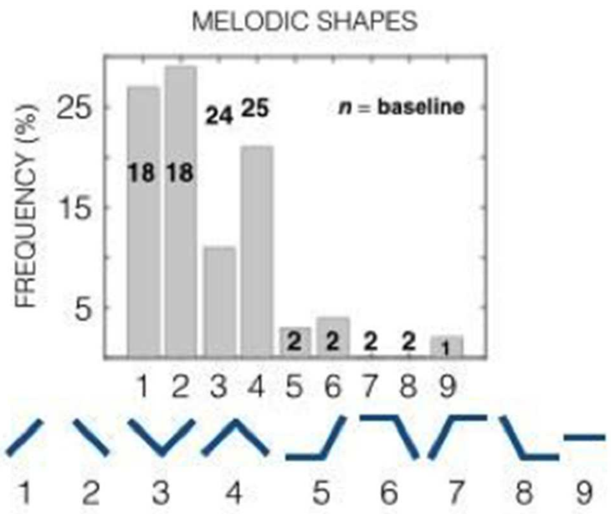
Supplementary Figure 3. Diagram-notation of the seeding material. Four seeds (S1-S4) of 5-tone sequences each, were used as starting material for the transmission chains. Each circle represents a single tone of the sequence. The blue line represents the changes in the melodic contour.



Supplementary Figure 4. Diagram-notations of melodic patterns produced in chain 1. Generations are ordered from top (seed or G1) to bottom. In each generation a set of 5 melodies was produced, corresponding to one of the 5 emotions. Each column represent the evolution of a single melodic code from G1 to G7.



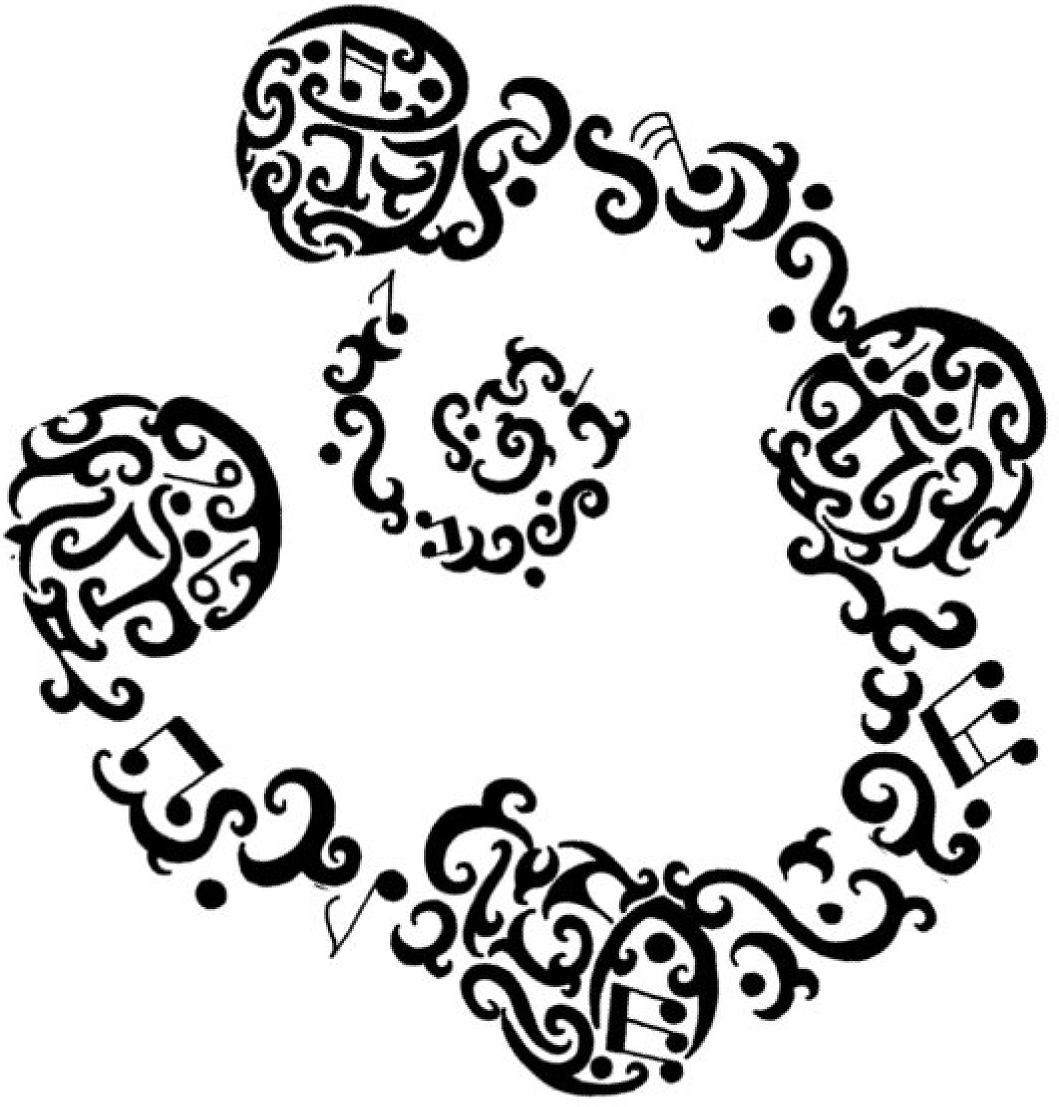
Supplementary Figure 5. Mean similarity between signals and relative (a) retrograde and (b) retrograde inversion transformations. No significant increase was observed between the first and the last generation. No differences with the baseline level were reported in the first and last generations.



Supplementary Figure 6. Distribution of string contour shape. Statistical distribution of 9 melodic contour configurations discussed by Tierney (2001). Bar height is equal to the mean frequency value computed across chains and generations. Baseline values, obtained by shuffling the original material (n=1000), are shown in bold.

Author contributions

Conceived and designed the experiments: ML and GB. Performed the experiments: ML. Analyzed the data: ML. Wrote the paper: ML and GB.



Chapter 3

Neural predictors of melodic regularization

Lumaca, M., & Baggio, G. (2016). Brain potentials predict learning, transmission and modification of an artificial symbolic system.
Social Cognitive and Affective Neuroscience, nsw112.

Abstract

It has recently been argued that symbolic systems evolve while they are being transmitted across generations of learners, gradually adapting to the relevant brain structures and processes. In the context of this hypothesis, little is known on whether individual differences in neural processing capacity account for aspects of variation observed in symbolic behavior and symbolic systems. We addressed this issue in the domain of auditory processing. We conducted a combined behavioral and EEG study on two successive days. On day 1, participants listened to standard and deviant 5-tone sequences: as in previous oddball studies, an MMN was elicited by deviant tones. On day 2, participants learned an artificial signaling system from a trained confederate of the experimenters in a coordination game in which 5-tone sequences were associated to affective meanings (emotion-laden pictures of human faces). In a subsequent game with identical structure, participants transmitted and occasionally changed the signaling system learned during the first game. The MMN latency from day 1 predicted learning, transmission and structural modification of signaling systems on day 2. Our study introduces neurophysiological methods into research on cultural transmission and evolution, and relates aspects of variation in symbolic systems to individual differences in neural information processing.

3.1. Introduction

Human symbolic systems such as language and music evolve while they are being acquired and transmitted by learners from subsequent generations. It has recently been hypothesized that the brain co-determines the direction of cultural evolution: symbol systems would gradually adapt to the constraints imposed by the relevant neural circuits, becoming easier to learn or process by inheriting specific structural and semantic properties (Christiansen & Chater, 2008; Dehaene & Cohen, 2007). This view has found some support in research on language (Kirby et al., 2008), arithmetic and reading (Dehaene, Cohen, Morais, & Kolinsky, 2015; Hannagan, Amedi, Cohen, Dehaene-Lambertz, & Dehaene, 2015), where it has been invoked to account for universal aspects of such systems. However, little is known about variation: would subtle yet possibly critical differences in brain structure and function across individuals account for some of the forms of variation (e.g., across individuals or over time) observed in cultural symbolic systems Dediu & Ladd, 2007).

Here we provide the first experimental evidence that symbol systems reflect individual cognitive processing capabilities when they are culturally learned and transmitted. We set out to test whether individual differences in neural information processing, as revealed by event-related potentials (ERPs), would predict how individuals learn, transmit and modify an artificial symbolic system. Our methodology involved two elements: (1) a laboratory model of cultural transmission, and (2) a neural marker of processing efficiency for structured sequences.

We used signaling games as a model of cultural learning and transmission (Lewis, 1969; Moreno & Baggio, 2015; Nowak & Baggio, 2016; Skyrms, 2010). In signaling games, the goal for the sender and receiver is to converge on a mapping of signals to meanings (or ‘code’) via iterated signaling rounds (see Methods). Our participants played 2 signaling games: first, as receivers, they learned a mapping of signals (5-tone sequences; the signal set) to meanings (simple and compound emotions); then, as senders, they transmitted the system learned in the first game (Fig. 2). Artificial languages are generally assumed to be relevant to understand natural languages. Similarly, artificial tone systems can be used to study actual musical systems. Moreover, music may be seen as a cultural symbolic system (Cross, 2005): it has often been used throughout its history to convey cognitive or affective meaning (Juslin & Sloboda, 2001). Our tone sequences were indeed endowed with affective meanings, referring to facial expressions of emotions: a link between facial expression and music-evoked emotions is supported by neural and behavioral research (Gordon et al., 2014; Hsieh et al., 2012; Palmer et al., 2013).

As a neural marker of auditory processing we used the mismatch negativity (MMN). The MMN is an ERP component peaking between 100 and 220 ms from the onset of a deviant event disrupting a standard sequence (Näätänen et al., 1978). Its elicitation therefore depends on whether a memory trace for the standard sequence has been formed (Alho, Sainio, Sajaniemi, Reinikainen, & Näätänen, 1990). Its latency and amplitude are correlated with pattern discrimination abilities in individuals (Näätänen et al., 1993). The MMN has been seen as a marker of processing efficiency (Gottselig et al., 2004; Tervaniemi et al., 2001), and as a means to investigate individual differences in auditory perception (Tervaniemi, Ilvonen, Karma, Alho, & Näätänen, 1997).

To test whether individual differences in auditory processing, as measured by MMN peak latencies and amplitudes, predict learning, transmission and structural modification of signaling systems, we designed two experimental sessions on two successive days: on day 1, participants underwent standard auditory MMN acquisition; on day 2, they played first as receivers (learners) then as senders (transmitters) in two successive signaling games (Methods, Study Design). The MMN stimuli used on day 1 and the signals used on day 2 were 5-tone sequences. If MMN amplitudes or latencies predict learning, transmission and reorganization of signaling systems, this would suggest a direct link between individual differences in neural processing and variation in cultural symbolic systems.

3.2. Methods

3.2.1. Participants

Twenty-one right-handed native speakers of Italian (9 female; mean age 23.71) with no hearing or neurological disorders participated in the study. Three (1 female) had an excess of movement artifacts in the EEG data; the remaining 18 were included in the final data analysis (mean age 23.71). To reduce expertise effects on auditory processing (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004; Seppänen, Brattico, & Tervaniemi, 2007), we included only musically naïve individuals.

3.2.2. Study design

Each subject participated in 2 sessions on 2 consecutive days (24h apart). The EEG session took place on day 1. At the start of the session, we assessed participants' working memory with a digit span test (Orsini et al., 1987) and a complex span tests (Foster et al., 2015). We recorded EEG data while tone sequences were delivered in 3 blocks with high-, low-, and 0-entropy sequences (see Stimuli) using loudspeakers (80 dB). During EEG recording, participants watched a silent subtitled movie in a 8×11 cm frame in the center of the screen. They were informed there would be questions about the movie at the end of the session, in order to divert their attention from the stimuli.

On day 2, the participant played 2 signaling games, first as receiver (Game 1), then as sender (Game 2) (see Transmission design; Fig. 2), with a confederate of the experimenters; sender and receiver roles were fixed. In Game 1, the participant (receiver) learned a signaling system where the signals were 5 5-tone sequences and the meanings were 3 basic and 2 compound emotions. In Game 2, the participant (sender) had to transmit the signaling system learned in Game 1, as they recalled it, to the receiver. The signaling games took place in a quiet room with two screens aligned back-to-back, with a curtain between them, preventing players from looking at or communicating with each other. Written instructions were given to the participant before the beginning of each session (Suppl. Inf. 1.1.1 and 1.1.2). Players used a full-size keyboard and headphones.

3.2.3. Bohlen-Pierce scale

Sequences were built using tones of the Bohlen-Pierce scale (Mathews et al., 1988). The equal tempered Bohlen-Pierce scale is based on 13 logarithmically even divisions of a tritave (3:1 frequency ratio). The frequency (F) of each note is: $F = k * 3^{(n/13)}$; k is the reference pitch frequency; n is the number of steps on the scale. This scale was used instead of the Western diatonic scale to prevent participants from exploiting prior intuitions about musical harmony or melody, which may affect perception of tone sequences. To avoid carry-over effects of the stimuli learned on day 1 (during the EEG session) on those used on day 2 in the signaling games, we set different n values to build the EEG stimuli and the materials used by the confederate in the signaling games (see below).

3.2.4. Day 1 - EEG experiment

3.2.4.1. Stimuli

Tone sequences of different complexity (high-entropy, low-entropy and 0-entropy or control) were presented in 3 blocks. In each block, 3 types of tone sequences were used: standards (80% frequency), contour deviants (C-deviant, 10%) and interval deviants (I-deviant, 10%); for control stimuli, a single deviant was used (Fig. 1). A session consisted of 8 sub-blocks with 250 stimuli each, and the control block included 4 sub-blocks. The order of the sessions (high-entropy, low-entropy and control) was counterbalanced across participants. The complexity of tone sequences was defined as the complexity of their contour transforms (contour entropy) where '+' stands for upward intervals, '-' for downward intervals, and '=' for repeated tones. We produced high-entropy standard stimuli (- + - +; Shannon Entropy = 1), low-entropy standard stimuli (+ + + -; SE = 0.81), and 0-entropy stimuli with repeated tones (= = = =; SE =

0). In contour deviants, the 4th tone violated the surface structure of standard sequences, but not interval size; vice versa for interval deviants on the 5th position. The deviant was always the same as the first tone in a sequence (Tervaniemi et al., 2001). Each sequence consisted of 5 50-ms sinusoidal tones from the frequencies 440, 521, 567, 617, 730.6 Hz (low register), separated by 50-ms of silence (interstimulus interval, ISI 750 ms). Sequences were transposed over trials at 3 registers of the Bohlen-Pierce scale (lowest frequencies: 440, 478, 567 Hz).

3.2.4.2. Data acquisition

The EEG was acquired using a Biosemi (Amsterdam, The Netherlands) ActiveTwo system with 128 active electrodes, digitized at 1024 Hz, off-line down sampled to 500 Hz and bandpass filtered (0.1-30 Hz). The EEG was referenced on-line to the average of signals from the left and right mastoid channels. The EOG was recorded by 2 electrodes above and below the right eye, and by 2 electrodes placed at the outer canthus of each eye. Data analysis was performed in ERPlab (Lopez-Calderon & Luck, 2014) and EEGlab (Delorme & Makeig, 2004). Epochs were time-locked to the onset of the critical (4th or 5th) tones for the standard and deviant stimuli, in a time window from -200 ms (baseline) to 800 ms after tone onset. Pre-processing included rejection of trials with ocular artifacts using a 150 ms step-function for the bipolar horizontal and vertical EOG channels (threshold at 35 µV; step 10 ms). A 200 ms-wide moving window was slid across the data in all 128 channels: epochs where the peak-to-peak offset exceeded 100 µV in any window were rejected. The remaining epochs were averaged separately for each condition (deviant or standard). Mean peak latencies and amplitudes of individual difference waves were calculated in each of 4 scalp quadrants (posterior, right, anterior, left; 32 channels each), and in each of 4 120-ms time windows (0-480 ms after tone onset), selected according to the timing characteristics of the MMN (100-220 ms from deviant tone onset; (Näätänen et al., 1978; Paavilainen, 2013) and after inspection of ERP data.

3.2.4.3. Data analysis

A mismatch negativity (MMN) is an enhanced negative ERP component in response to deviants relative to standard stimuli, larger over frontal sites at 100-220 ms from stimulus onset (Näätänen et al., 1978; Paavilainen, 2013). We used a four-way repeated-measures ANOVA with mean MMN amplitude as dependent variable in each of 3 conditions, and 4 factors: Stimulus Type (standard or deviant), Temporal Window (4 levels, as above), Quadrant (4 levels, as above) and, nested on the latter, Electrode (32*4 levels), using the Greenhouse-Geisser correction when appropriate. Wilcoxon signed-rank tests were used in post-hoc analyses.

To investigate whether individual differences in auditory processing predict learning, transmission and structural modification of signaling systems, we first measured MMN amplitudes and peak latencies in individual difference waves (deviant minus standard stimuli). Next, we computed correlations between MMN peak latencies (or mean amplitudes) in each participant with various measures of behavior in signaling games (see below). P-values from the correlation tests were Bonferroni corrected at $\alpha = 0.05/12 = 0.004$; 12 is the number of times each independent variable (amplitude or peak latency) is correlated with behavioral measures.

3.2.5. Day 2 - Signaling games

3.2.5.1. Signaling games

Signaling games are games of incomplete information in which a sender and a receiver develop a mapping of signals to meanings via repeated interactions (Lewis, 1969; Skyrms, 2010). In the standard form, the sender privately witnesses an event or state and uses a signal to inform the receiver on the identity of that state. Upon receiving the signal, the receiver takes appropriate action to match the state's identity. Communication may succeed or fail depending on whether the state and action match or not. The players receive feedback as to whether the trial was successful. (Moreno & Baggio, 2015) showed that when sender and receiver roles are fixed within a game, the mapping is largely determined by the sender and learned by the receiver. In these conditions, signaling games are a viable laboratory model of social learning and cultural transmission of artificial symbol systems (for replications and extensions, see Nowak & Baggio, 2016). In our experiment signals were constituted by 5 sine-wave tones (200 ms length; inter-tone interval 50 ms). In the Bohlen-Pierce equation, we defined $k = 440$ Hz and $n = 0, 4, 6, 7$, or 10 to obtain the maximum number of tone combinations with a low-integer frequency ratio. The states were 5 basic (joy, peace or sadness) or compound emotions (peace+joy or peace+sadness), depicted as facial expressions. We recruited 4 professional actors (2 male, 2 female) for a photo shoot session: they were instructed to express naturally each basic emotion; one photo per emotion per actor was included in the final stimulus set. Facial expressions for compound emotions were generated by dividing up expressions of basic emotions into two regions (Ekman & Friesen, 2003): the upper and lower face. Compound emotions were produced by merging the upper face portion of peace and the lower face portion of joy (peace+joy) or sadness (peace+sadness).

3.2.5.2. Transmission design

Fig. 2 (right panel) shows the structure of signaling games. In Game 1, the participant played as receiver and the confederate played as sender. Before the start of the game, the confederate was trained to deliver one of 3 sets of stimuli. In each melodic set, signals with different contour complexity were included: 2 high entropy (Shannon entropy, Entropy=1), 2 low entropy (Entropy=0.81), 1 monotone stimulus (control; Entropy=0). Each set was randomly assigned to an equal number (7) of participants. We expected the participant to learn the associations between tone sequences and emotions consistently used by the confederate in Game 1. In Game 2, the participant played as sender and the confederate as receiver. The sender was asked to reproduce and transmit the system of tone sequences learned in the previous game as they remembered it.

3.2.5.3. Procedure

Fig. 2 (left panel) shows a single trial in signaling games. The sender is privately shown a facial expression from an actor’s set (e.g., joy; 3 sec duration). Next, he is expected to generate an isochronous 5-tone signal from a pool of 5 tones using a standard keyboard, to inform the receiver on the state the sender has seen. The relation between keyboard keys and tones was randomized over trials, so in each trial the sender had to learn anew the key-to-tone mappings[2]. Unheard by the receiver, the sender could try out each composition of the 5-tone signal at will; the relation between keyboard keys and tones was fixed within a single trial. Visual cues were provided on the screen showing the ‘height’ of each tone. The signal was sent to the receiver, who listened to it via headphones, and chose, using the numeric keypad, one of the 5 facial expressions. All 5 expressions were displayed simultaneously on the screen; their position was reshuffled randomly for each trial. Feedback (3 sec duration) was given to both players simultaneously: the left-hand side of the screen showed the face seen by the sender, and the right-hand side showed the face chosen by the receiver. For positive feedback, both faces were displayed in a green frame, while for negative feedback a red frame was used. Participants had to play exactly 70 trials in Game 1 and 50 trials in Game 2. Each state (emotion) occurred equally frequently. There was no reward for speed in a trial or in the game: coordination was the only reward for participants.

3.2.5.4. Data analysis

Our aim was to determine whether MMN peak latencies or amplitudes predict individual performance and changes in the organization of tone sequences in signaling games. The simplicity and formal explicitness of signaling games allows us to define a

---

[2] A partial exception here was the confederate input in Game 1, see Suppl. Inf. 1.2.1.

set of intuitive, rigorous measures of similarity between two tone sequences (Hamming distance), learning or recall (coordination, transmission, innovation), direction of transmission from sender to receiver or vice versa (asymmetry), and structural modification (entropy and proximity) (Suppl. Fig. 2).

3.2.5.5. Signaling measures

Measures of learning and recall were computed using a measure of similarity, the inverse of the Hamming distance (Hamming, 1986). The elements compared were the 5 tones (tone distance) or interval directions (ups and downs) of adjacent tones (contour distance). A normalized output to [0 1] was given by S/L: S is the number of shared items and L the length of a sequence. Values close to 1 are associated to similar strings. Using the Hamming distance, we measured to what extent the signaling system transmitted by the confederate was actually learned by the participant (receiver) in Game 1 (coordination), and to what extent the participant (sender) then reproduced (transmission) or modified (innovation) that system in Game 2. Formally, coordination is the extent to which sender and receiver during Game 1 use the same mapping of states to signals (i.e., the extent to which they share a signaling system), and is defined as the mean similarity between the signal used by the sender for a given emotion and the set of signals mapped by the participant to the same emotion during the second half of Game 1, where player responses are more consistent and show fewer errors. It ranges from 1 (shared signaling system; the participant uses the same mapping of states to signals as the confederate, and has therefore learned that mapping) to 0 (different signals for every emotion). Transmission was measured as the contour similarity between pairs of tone sequences for a particular emotion produced with greater frequency by the senders in Game 1 (confederate) and Game 2 (participant) (between-player measure). Faithful transmission of signaling systems learned in Game 1 yields values close to 1. Innovation is the Hamming contour distance between the signals acquired during Game 1 and transmitted by the participant during Game 2 denoting a particular emotion (within-player measure). Values close to 1 indicate that a significant number of changes were introduced in the signaling system by the participant. If coordination in Game 1 is imperfect (the signaling system transmitted by the confederate and the signaling system learned by the participant do not exactly correspond), transmission and innovation do not mirror each other. A measure of asymmetry during coordination was given by subtracting the number of changes introduced in the signaling system by the sender from the number of changes made by the receiver, divided by the total number of changes (Moreno & Baggio, 2015); it ranges from 1 (the sender is modifying the signaling system more frequently) to -1 (vice versa).

Entropy

The Shannon Entropy (Shannon 1949) of tone sequences (and of the relative contour transforms) was used as measure of the complexity of melodic surface structures:

$H(X) = -\Sigma P(xi)\log 2P(xi)$

X is the string of ups and downs (contour smoothness) or tones (sequence complexity) and P(xi) is the probability to find item xi in that string position. A single mean entropy value was computed for the entire set of sequences, and separately for the relative contour transforms. The absolute difference in mean entropy values was computed between the two games and used in the correlation analysis with MMN latencies/amplitudes.

Proximity

Measures of perceptual proximity were applied to the sequences received and reproduced by the participant (interval distribution, interval compression ratio). The distribution of unison (0 macrotones), small (1 or 2 macrotones) and large intervals (3, 4, 6, 7 or 10 macrotones) was calculated for each set of melodies. The absolute difference in the frequency of unison, small or large intervals was computed between the two games and was used in the correlation analysis with MMN latencies or amplitudes. The interval compression ratio (Tierney et al., 2011) represents a bias in tone sequences toward smaller intervals. The interval compression ratio is defined as the mean absolute interval size (in macrotones) of a tone sequence randomly shuffled (n=100), divided by the mean absolute interval size of the original sequence. Because a scrambled version of a proximal melody typically has larger intervals than the original (Von Hippel, 2000), larger values for this ratio are associated to compressed, small-interval melodies.

3.3. Results

3.3.1. Mismatch negativity

The time-course and topographical distribution of the effect elicited by deviant stimuli are typical of the MMN (Figs. 3-4; Suppl. Figs. 2-4), with a frontal maximum between 120 and 240 ms from deviant onset. Statistical tests confirmed an MMN effect in all three conditions (Table 1). Interaction effects support a stronger MMN at 120-240 ms in the frontal quadrant.

3.3.2. Asymmetry and coordination

Asymmetry in Game 1 was -1 by design (the confederate never changed the current signaling system), and in Game 2 it was negative and different from 0 (median=-0.12; n=18; Z=3.48; p<.001): the sender (participant) changed his mapping of states to signals (learned in Game 1, possibly with modifications) less frequently than the receiver (confederate). Coordination was different from 0 (median=0.77; n=90, Z=0.86, p<.001): the tone systems were effectively learned during the game. These data show a net flow of information from the sender to the receiver in both games.

3.3.3. Transmission and innovation

Transmission of tone sequences was faithful and significantly different from 0 (contour distance, median=0.75; n=90, Z=7.61, p<.001), showing scarce innovation (median=0.25; n=90, Z=6.74, p<.001). Using tone distance, instead, we observed lower levels of transmission (median=0.2) and higher innovation (median=0.8), reflecting ease of recall and production of the contours of sequences. This justifies our use of contour distance as a measure of transmission and innovation in subsequent analyses.

3.3.4. Structural changes

Tone sequences were restructured in the course of the signaling games. The entropy of tone sequences decreased from Game 1 to 2 (Z=-2.673, p=0.008), more than contour entropy (Z=-1.835, p=0.06). Both these changes reflect the emergence of smoother sequences: tones were repeated or varied in small steps. Median interval compression ratio values (Z=-2.577, p=0.01) indicate changes toward smaller intervals.

To examine whether ERPs predict participant behavior in signaling games, we correlated MMN latencies or amplitudes with behavioral performance or structural changes to tone sequences during Game 2. Below we focus on the latency (120-240 ms) typical of the MMN (Suppl. Tabs. 1-8).

3.3.5. Neural predictors of learning and transmission

Fig. 5 (upper panel) shows behavioral measures (coordination, transmission, innovation accuracy) against MMN peak latencies in the anterior quadrant. MMN data from this quadrant only, in the high-entropy condition, were predictive of individual behavior (Suppl. Tab. 1). Peak latency was negatively correlated with coordination (p<0.004) and transmission (p<0.01): individuals with shorter MMN peak latencies learned more efficiently the signaling systems in Game 1, and transmitted them more faithfully in Game 2. Shorter MMN latencies also predicted fewer changes to the signaling system across games (innovation, p<0.01). Moreover, peak latencies from the anterior quadrant predicted accuracy in Game 1 ($p<0.004$). In Game 2, we observed a positive correlation with the number of changes introduced in the signaling system before coordination occurred, which suggests a greater effort to coordinate for participants with longer MMN peak latencies (Suppl. Tab. 3). Together, these results point to a relation between longer MMN peak latencies and less accurate or less faithful acquisition and transmission of tonal material.

No correlations in the remaining conditions (and quadrant) were significant (Suppl. Tabs. 1, 3). Using mean MMN amplitudes, correlations were weaker and right lateralized ($p<0.05$; Suppl. Tabs. 2, 4). No correlations were found in the 240-360 ms post-MMN window (Suppl. Tabs. 5-8).

3.3.6. Neural predictors of structural changes

Fig. 5 (lower panel) shows absolute changes in tonal (contour smoothness, sequence complexity) and interval structure (percentage of small and large intervals) against peak latencies in the frontal scalp quadrant. Positive correlations were found with contour smoothness ($p<0.05$), sequence complexity ($p<.001$) and changes in the percentage of smaller ($p=0.01$) and larger intervals ($p=0.004$). These data reflect a reorganization of the melodic surface and interval distribution of sequences, mainly introduced by individuals with longer MMN latencies. In contrast, individuals with shorter MMN latencies were more conservative: specifically, they tended to maintain the melodic structures acquired in the first game.

3.3.7. Working memory

Working memory span scores were not predictive of behavior or structural changes in signaling games (all p-values>0.05).

3.4. Discussion

The main finding of the present study is that inter-individual variability in pre-attentive processing of tone sequences, as revealed by MMN latencies, predicts subsequent learning, transmission and structural reorganization of signaling systems by participants. Below we discuss the possible origins and implications of this variability in the context of experimental research on cultural transmission.

3.4.1. Individual differences in MMN latencies

The auditory MMN reflects pre-attentive sensitivity to regularities and how accurately these are encoded as auditory traces (Alain, Achim, & Woods, 1999; Alain, Cortese, et al., 1999). Can the MNN be used to track individual differences in auditory processing capacity? Previous studies have shown correlations between the MNN characteristics and individual discrimination capability with complex spectrotemporal patterns (Näätänen et al., 1993; Tervaniemi et al., 2001). Näätänen et al. (1993) showed that poor auditory discrimination is related to a smaller MMN: a monotonic increase in MMN amplitude was observed for participants improving their performance over time. Converging results were obtained with hit-rates and reaction times (Kraus et al., 1996; Menning, Roberts, & Pantev, 2000; Tiitinen, May, Reinikainen, & Näätänen, 1994; Tremblay, Kraus, & McGee, 1998) and fMRI (Zatorre, 2013): e.g., auditory cortex responses to microtonal intervals (smaller than the Western semitone) predicted individual discrimination of microtonal patterns (Zatorre, Delhommeau, & Zarate, 2012). Overall, these data point to neural predispositions (not necessarily innate) in some individuals for processing complex auditory objects (Zatorre, 2013).

One possibility is that ERP latencies reflect connectivity strengths in the relevant networks, while amplitudes are linked to the extent of active tissue (Baggio, 2012; Baggio & Hagoort, 2011; Cardenas et al., 2005). This might explain why in our experiment the best predictors were MMN latencies and not amplitudes. Variation in MMN latencies may then be related to learning or maturational states of the relevant cortical networks (Choudhury & Benasich, 2011; Moore & Guan, 2001; Trainor et al., 2003). Inter-individual differences in temporal integration, moreover, may be a source of variation in MMN latencies: temporal integration is used by the auditory system to track input regularities, create a unified perceptual model (Bregman & Ahad, 1995) and predict upcoming events. Violations of expected regularities would then explain the generation of mismatch-related brain potentials (Winkler, 2007; Winkler, Denham, et al., 2009).

Temporal integration is key to create and maintain auditory representations of the environment (Sussman, Ritter, & Vaughan, 1999). In people with poor integration capacity, rapidly presented information, as in our experiment, may limit the accuracy of the perceptual model, and delay detection of the deviant sound. Sensory memory might be rapidly overwritten by novel inputs unless the stimuli are not rapidly processed (Christiansen & Chater, 2016). Although impairments in rapid temporal integration mechanisms have been reported in special or clinical populations (Bishop, 2007; Williamson & Stewart, 2010), it is likely that temporal integration abilities occur in a continuum. Individual differences in MMN characteristics are indeed more pronounced between- than within-individuals (Pekkonen, Rinne, & Näätänen, 1995). Collecting these evidences, we suggest that interindividual differences in MMN latencies observed here reveal individual constraints in auditory processing, rather than contextual, or global, non-functional factors. In addition, if these low-level mechanisms govern perceptual information processing (Näätänen & Winkler, 1999), they may account for performance at the behavioral level (Novak, Ritter, & Vaughan, 1992; Tervaniemi, Ilvonen, et al., 1997; Tiitinen et al., 1994).

3.4.2. Sequence complexity and behavioral prediction

Demands on temporal integration change with the complexity of the rules violated (Todd, Myers, Pirillo, & Drysdale, 2010) as well as with individual differences (Näätänen & Winkler, 1999). From low- to high entropy sequences, rules become more complex and less predictable, increasingly taxing the auditory system. Thus, control and low entropy melodies (Fujioka et al., 2004; Näätänen & Winkler, 1999; van Zuijen, Sussman, Winkler, Näätänen, & Tervaniemi, 2004) draw on primitive grouping mechanisms and ‘top-down schemata’ largely shared across individuals of the same musical culture (Snyder, 2008). Here, differences in temporal integration, reflected by MMNs, are minimal (van Zuijen et al., 2004). Instead, greater individual variation in ERPs has been shown for complex patterns (Boh, Herholz, Lappe, & Pantev, 2011; Näätänen et al., 1993). In this condition, different integration capacities may be reflected at the behavioral level (Näätänen & Winkler, 1999). Consistent with this notion, only MMN latencies recorded with high-entropy stimuli in our study were predictive of behavioral performance.

3.4.3. Learning, transmission and modification of symbolic systems

One key finding here is that learning and transmission are predicted by the timing of brain processes: shorter MMN latencies were observed in faithful transmitters; longer latencies were found in individuals who innovate more the artificial tone system. This agrees with previous findings (Näätänen et al., 1993; Schröger, Näätänen, & Paavilainen, 1992; Tervaniemi, Schröger, & Näätänen, 1997). The result that latencies correlate with learning (coordination) and successive recall (transmission) provide additional support to the memory interpretation of the MMN, up to now addressed using auditory discrimination tasks (Näätänen, Paavilainen, Rinne, & Alho, 2007).

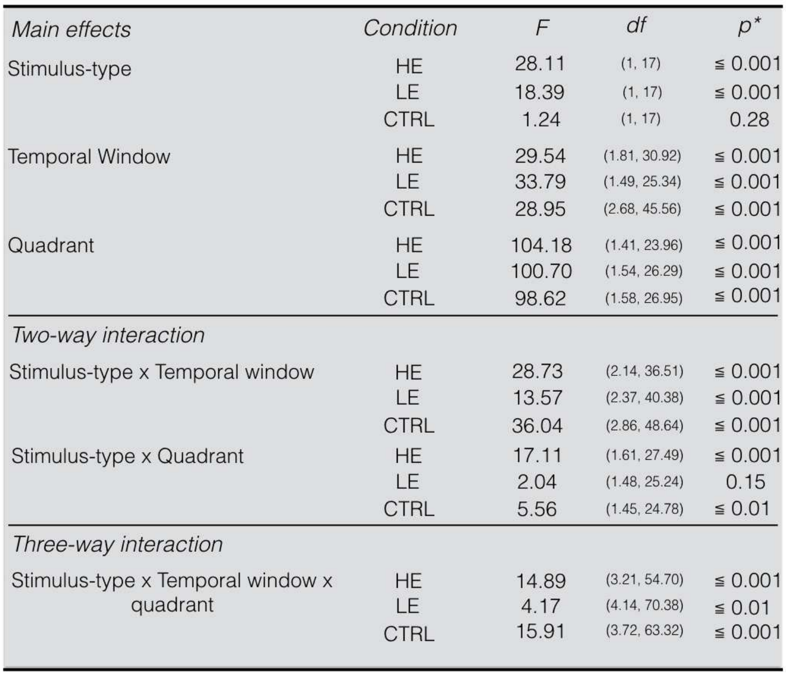
Perhaps most interestingly, MMN latencies predict changes in the structure of signals: individuals with longer latencies reorganized the melodic contour and interval distributions of sequences more often; smoother and proximal melodic forms became more frequent from Game 1 to Game 2. Importantly, these changes follow the Gestalt principles of auditory perceptual proximity and good continuation: strings with such properties are easier to memorize and recall (Deutsch, 2012). Previous experimental studies support the view that cultural systems tend toward compressibility and simplicity (Tamariz & Kirby, 2015; Verhoef et al., 2013) to fit the capability and limits of human memory (Chater & Vitányi, 2003). Our work shows that these pressures can appear after a single round of learning and transmission, and suggests that the extent of the adaptation of cultural systems toward simpler forms may relate to individual constraints on information processing.

3.4.4. Individual processing constraints and cultural evolution

One leading hypothesis in cultural transmission research is that symbolic systems are shaped and constrained by properties of the human mind/brain (Christiansen & Chater, 2008). Humans transform information they receive from others, introducing features which fit general constraints on learning and memory (Boyd & Richerson, 1985). These constraints may be amplified during transmission, making the cultural system easier to acquire, use and transmit (Griffiths, Kalish, & Lewandowsky, 2008; Kirby, 2001; Kirby et al., 2008; Smith & Kirby, 2008; Verhoef, 2012). Their existence has so far been inferred largely from observations at the population level (Griffiths et al., 2008; Kirby et al., 2008). Such constraints are often assumed to be too small to be observed and measured in single individuals and, as a result, no evidence has supported this view by direct neural investigations.

Here, we have shown that such constraints can be manifest at the individual level after only one instance of transmission, leaving distinct behavioral and neural traces. We have shown that the extent of change varies significantly across individuals. Our study provides initial neural evidence for the role of learning and memory constraints in the cultural transmission of symbolic systems (Christiansen & Chater, 2008), and musical systems in particular (Huron, 2001; Kleeman, 1985; Sawa, 2002; Trainor, 2015).

3.5. Figures and tables



Main effects	*Condition*	*F*	*df*	*p**
Stimulus-type	HE	28.11	(1, 17)	≤ 0.001
	LE	18.39	(1, 17)	≤ 0.001
	CTRL	1.24	(1, 17)	0.28
Temporal Window	HE	29.54	(1.81, 30.92)	≤ 0.001
	LE	33.79	(1.49, 25.34)	≤ 0.001
	CTRL	28.95	(2.68, 45.56)	≤ 0.001
Quadrant	HE	104.18	(1.41, 23.96)	≤ 0.001
	LE	100.70	(1.54, 26.29)	≤ 0.001
	CTRL	98.62	(1.58, 26.95)	≤ 0.001
Two-way interaction				
Stimulus-type x Temporal window	HE	28.73	(2.14, 36.51)	≤ 0.001
	LE	13.57	(2.37, 40.38)	≤ 0.001
	CTRL	36.04	(2.86, 48.64)	≤ 0.001
Stimulus-type x Quadrant	HE	17.11	(1.61, 27.49)	≤ 0.001
	LE	2.04	(1.48, 25.24)	0.15
	CTRL	5.56	(1.45, 24.78)	≤ 0.01
Three-way interaction				
Stimulus-type x Temporal window x quadrant	HE	14.89	(3.21, 54.70)	≤ 0.001
	LE	4.17	(4.14, 70.38)	≤ 0.01
	CTRL	15.91	(3.72, 63.32)	≤ 0.001

* *p-values are Grenhouse-Geisser corrected*

Table 1. Results of ANOVA statistics on mean MMN amplitude values in the high entropy (HE), low entropy (LE) and control conditions (CTRL).

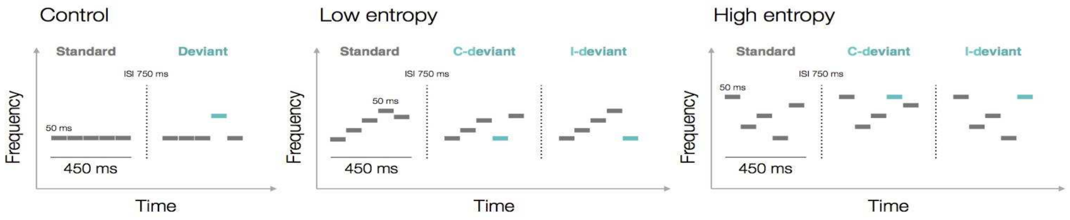


Fig. 1. Schematic illustration of the tone sequences presented to participants on day 1 of the experiment. The EEG was recorded while participants listened to these sequences. Each sequence consisted of 5 tones with 50 ms duration and 50 ms inter-tone intervals, i.e., 450 ms sequence duration. The interstimulus interval (ISI; the interval between the offset of the last tone in one sequence and the onset of the first tone in the next sequence) was 750 ms. Standard sequences of different complexity were used: control (monotone) sequences, low and high entropy sequences. Deviants were constructed such that the 4th tone

violated the melodic contour of the standard sequence (contour deviant or C-deviant) or the 5th tone violated the interval between the 4th and 5th tones in standard sequences while preserving melodic contour (interval deviant or I-deviant). The proportion of standards to deviants was 4:1. A blocked design was used, where the blocks were randomized across participants.

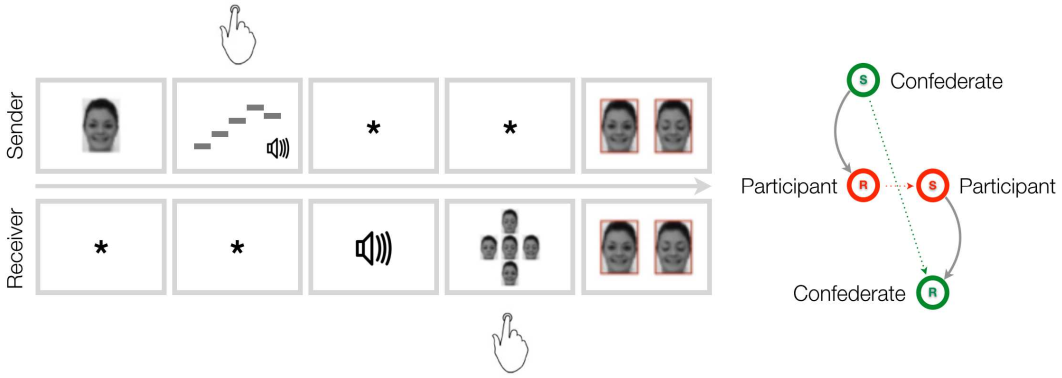


Fig. 2. Example of a trial from the signaling games (left panel) played by participants on day 2 of the study. The top and bottom rows show what the sender and receiver saw on their screens, respectively. The task for the sender was to compose a 5-tone sequence to be used as a signal of the simple or compound emotion expressed by the face presented on the screen at the start of the trial, and for the receiver to respond to that signal by choosing the face that the sender had seen. The sender and the receiver converged over trials on a shared mapping of signals (tone sequences) to meanings (emotions). Hand symbols indicate when the sender and the receiver had to produce a response. Feedback was given to both players simultaneously, displaying the face seen by the sender and the face selected by the receiver in a green frame (matching faces; correct) or in a red frame (mismatching faces; incorrect). Time flows from left to right. A diagram of the signaling games played by the participant is also shown (right panel). The participant played as receiver (R) with a confederate of the experimenters playing as sender (S) in Game 1, and roles switched in Game 2.

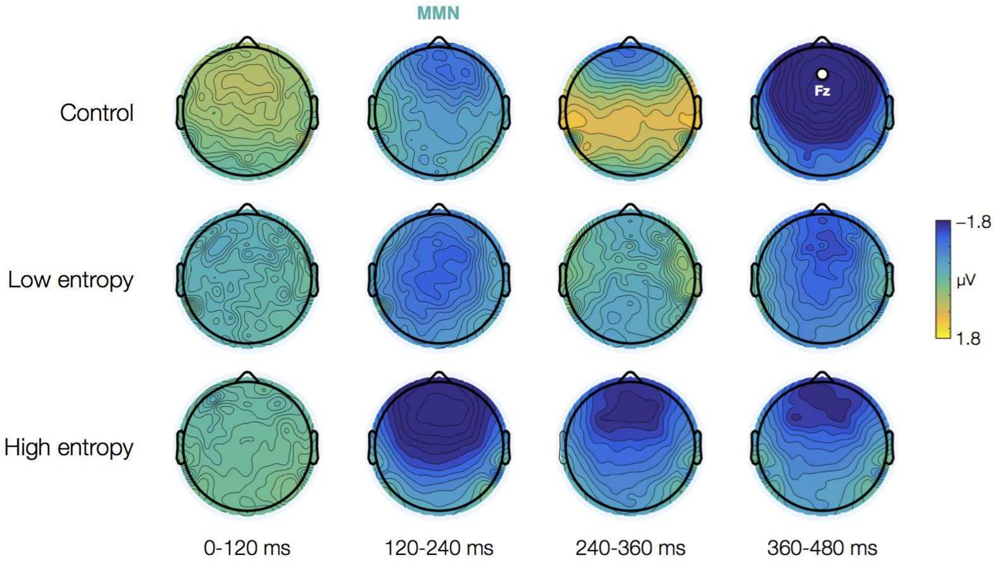


Fig. 3. Topographic isovoltage maps of grand-average (n=18) difference waves between ERPs to deviants and standards in control, low and high entropy sequences. The mismatch negativity (MMN) corresponds to the topographic maps shown at 120-240 ms; 0 ms is the onset of the deviant tone (4th or 5th) in a sequence.

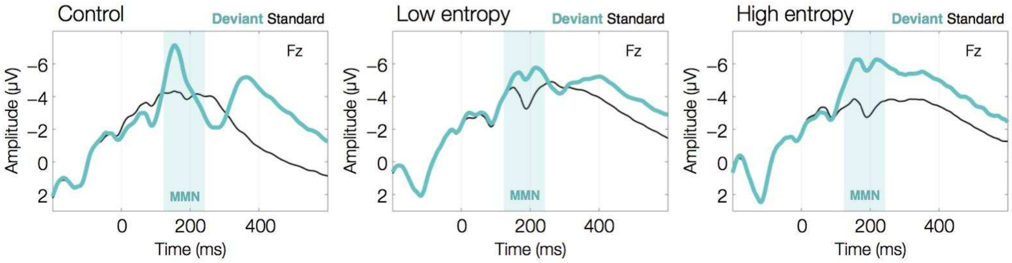


Fig. 4. Grand-average ERPs (n=18) from the frontal midline electrode (Fz) in response to the 4th and 5th tones in standard and deviant stimuli in control, low and high entropy sequences. The waveforms shown here were low-pass filtered at 30Hz.

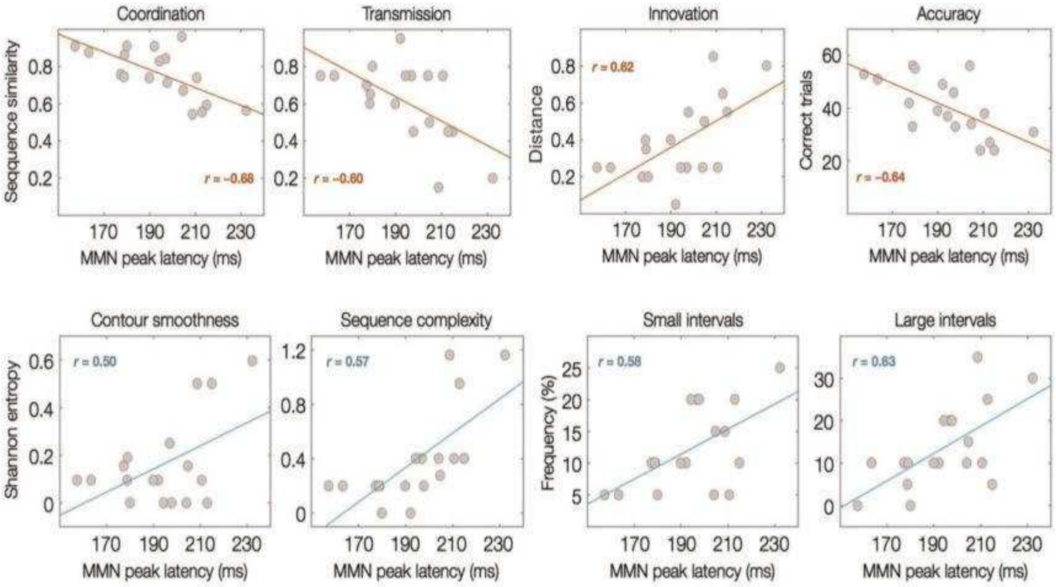


Fig. 5. Pearson product-moment correlations (r) between MMN peak latencies relative to the onset of the deviant tone (0 ms) in high entropy sequences recorded from the frontal scalp quadrant on day 1 (x-axis), and behavioral (top, orange) and structural (bottom, blue) measures from signaling games on day 2 (y-axis). All MMN peaks fall within the 120-240 ms window. See Methods and Suppl. Fig. 1 for further details on the measures. Each point on a scatterplot is one participant (n=18).

3.6. Supplementary material

1.1. Written instructions to participants

1.1.1. ***Day 1.*** During the EEG session, the participant was asked to watch a silent movie without paying any attention to the tone sequences played through the loudspeakers; the movie was subtitled in Italian. To direct their attention away from the tone sequences, we told participants that there would be questions about the movie at the end of the session.

1.1.2. ***Day 2.*** Prior to the signaling games (SGs) session, the participant was informed he would take part in two successive games, each played in a different role. In Game 1, he would be the receiver. He was asked to try to learn which facial expressions (emotions) corresponded to the 5-tone sequences sent by the other player (confederate). There was no reward for speed, either in a trial or in the game. Successful coordination was the only reward for participants. In Game 2, the participant played as sender. He was instructed to transmit the tone system acquired in Game 1, as he remembered it, to the receiver.

1.2. Confederate behavior

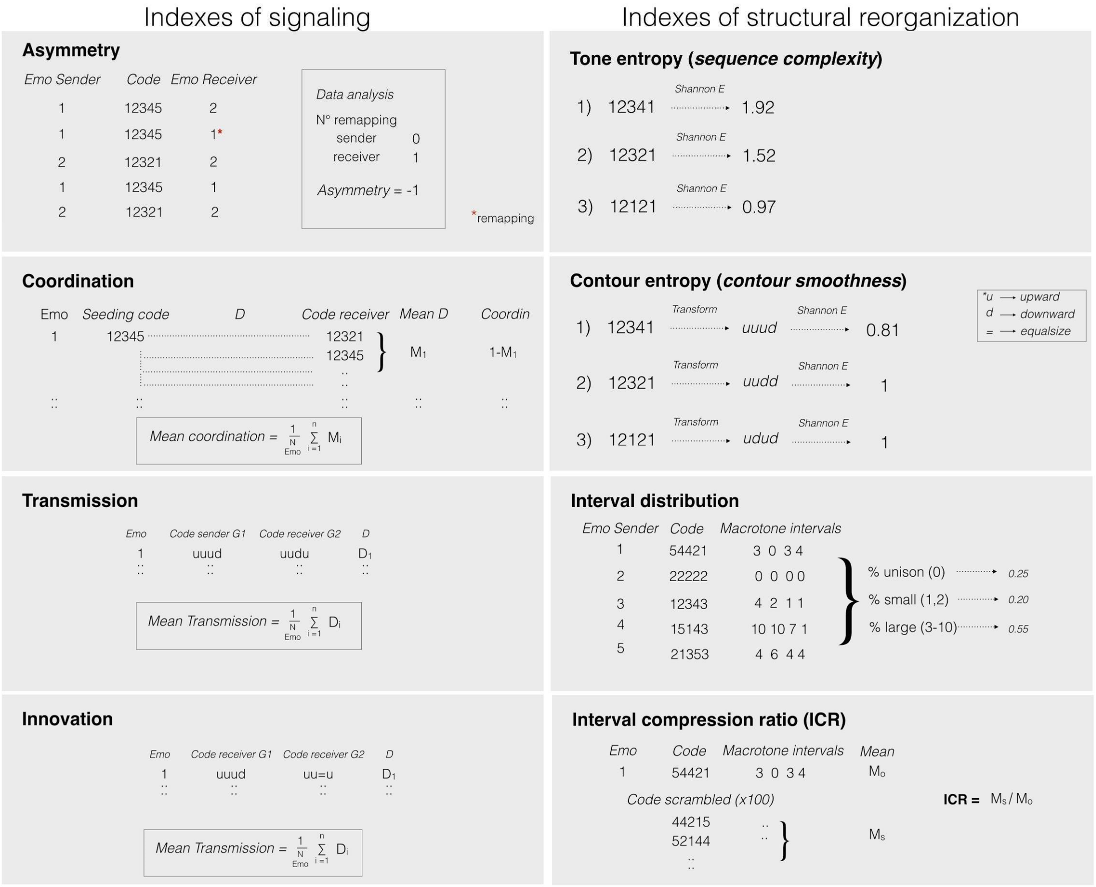
1.2.1. ***Game 1***. The confederate was trained on a particular mapping of emotions to keys on the keyboard (instead of tones): in Game 1, he played with no sound feedback (volume set at 0), to prevent him from altering the tone system based on auditory principles or prior knowledge, and from using (in any way) the ‘seed’ system in Game 2. The confederate was instructed to use consistently that mapping throughout Game 1, regardless of what the participant (receiver) would do.

1.2.2. ***Game 2.*** The confederate played according to the same written instructions given to the participant (receiver) in Game 1, namely, he was instructed to try to learn the tone system as produced by the participant (sender).

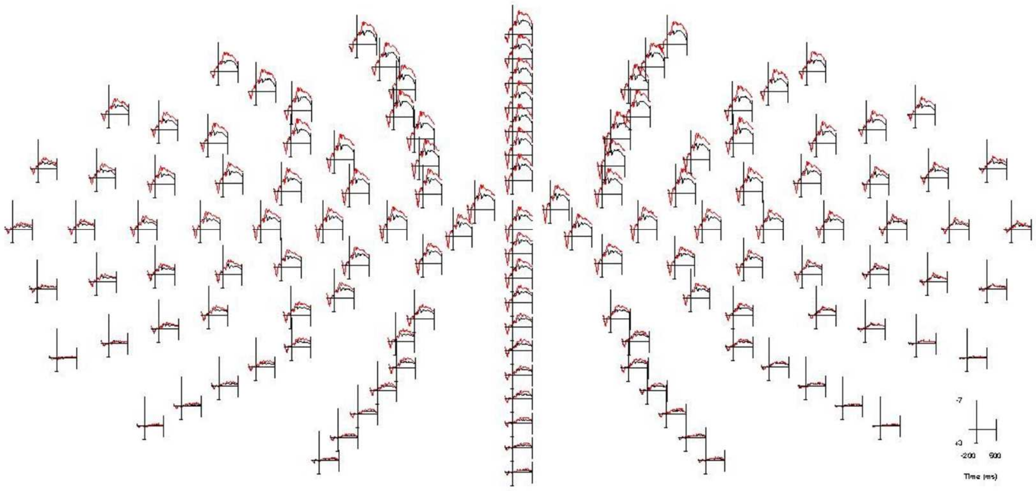
1.3. Ethics

The experiment was conducted according to the principles of the Helsinki declaration on research involving human subjects. Participants signed a written informed consent sheet to participate in the study. Data were anonymized upon collection by removing personal information. All procedures were approved by the Ethics Committee of SISSA.

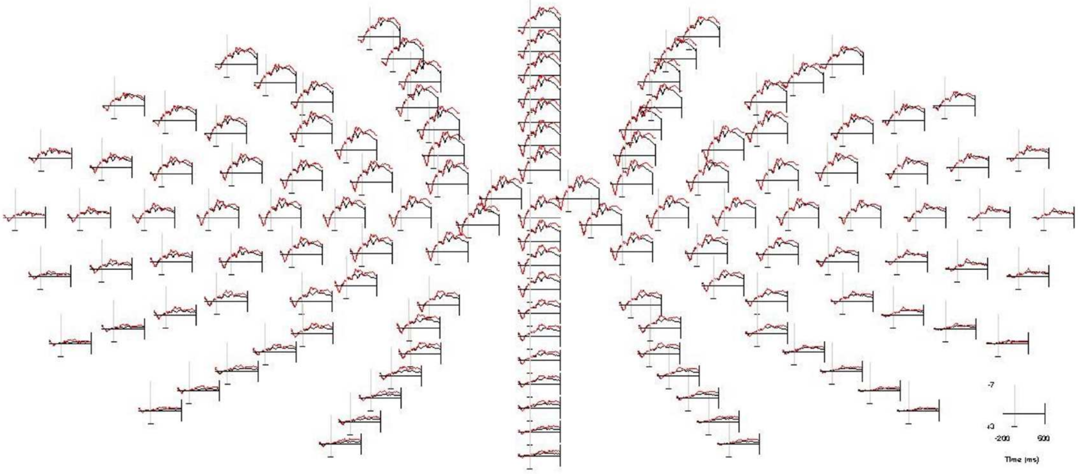
2 Supplementary figures



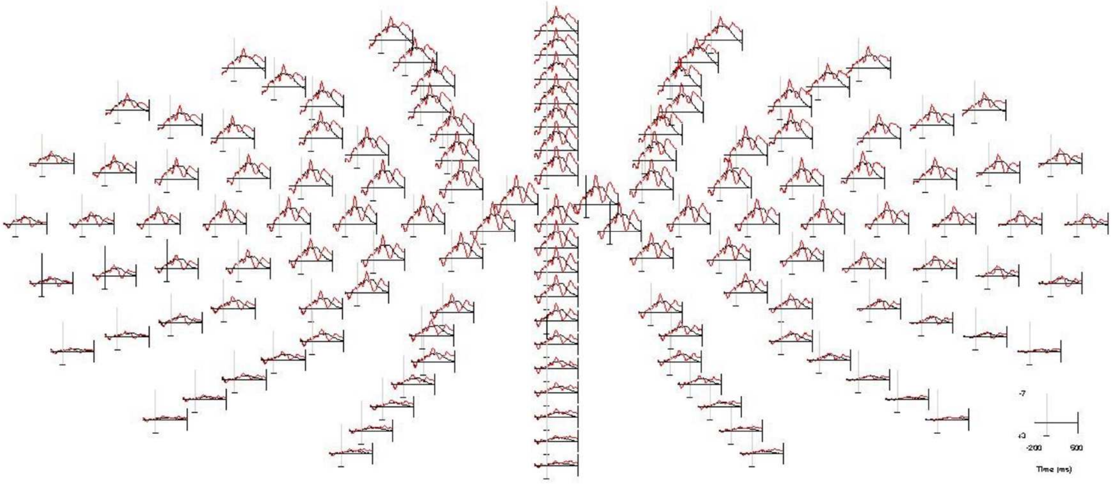
Supplementary Figure 1. Examples of the formal measures used in the structural analyses of tone sequences (for additional information, see Methods in the main article).



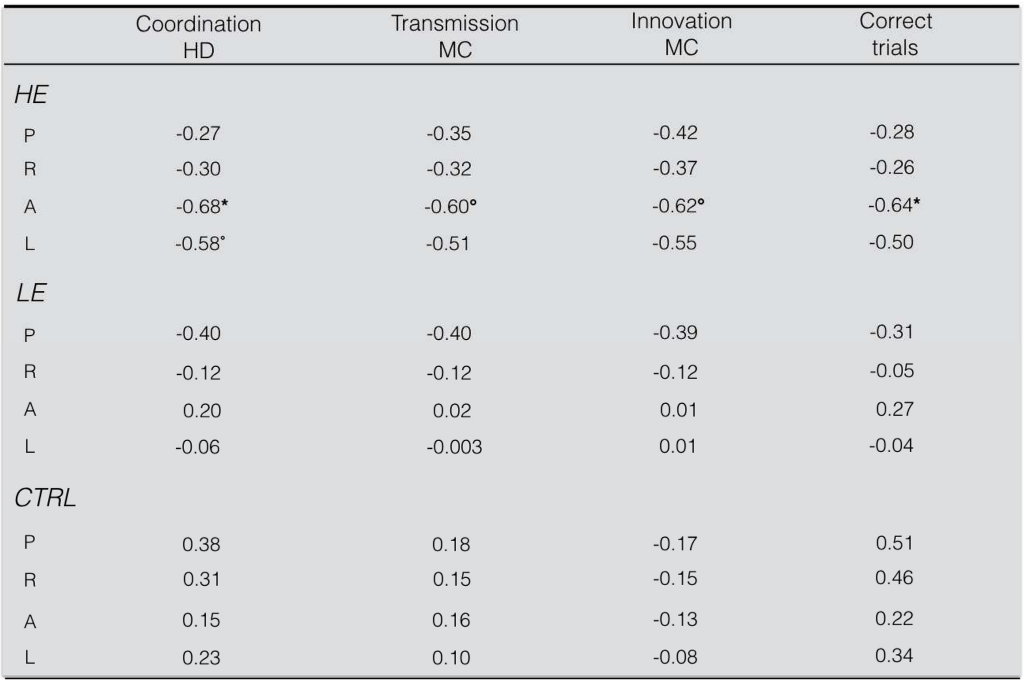
Supplementary Figure 2. Topographically-arranged plots (n=18) of standard (black) and deviant (red) ERP waveforms in the ***high entropy*** condition (all 128 channels). The onset of the critical tone (0 ms) is shown as a black vertical bar in each plot. Negative voltages are plotted upward. An MMN is visible at about 200 ms from the deviant tone onset.



Supplementary Figure 3. Topographically-arranged plots (n=18) of standard (black) and deviant (red) ERP waveforms in the ***low entropy*** condition (all 128 channels). The onset of the critical tone (0 ms) is shown as a black vertical bar in each plot. Negative voltages are plotted upward. An MMN is visible at about 200 ms from the deviant tone onset.



Supplementary Figure 4. Topographically-arranged plots (n=18) of standard (black) and deviant (red) ERP waveforms in the ***control*** condition (all 128 channels). The onset of the critical tone (0 ms) is shown as a black vertical bar in each plot. Negative voltages are plotted upward. An MMN is visible at about 200 ms from the deviant tone onset.



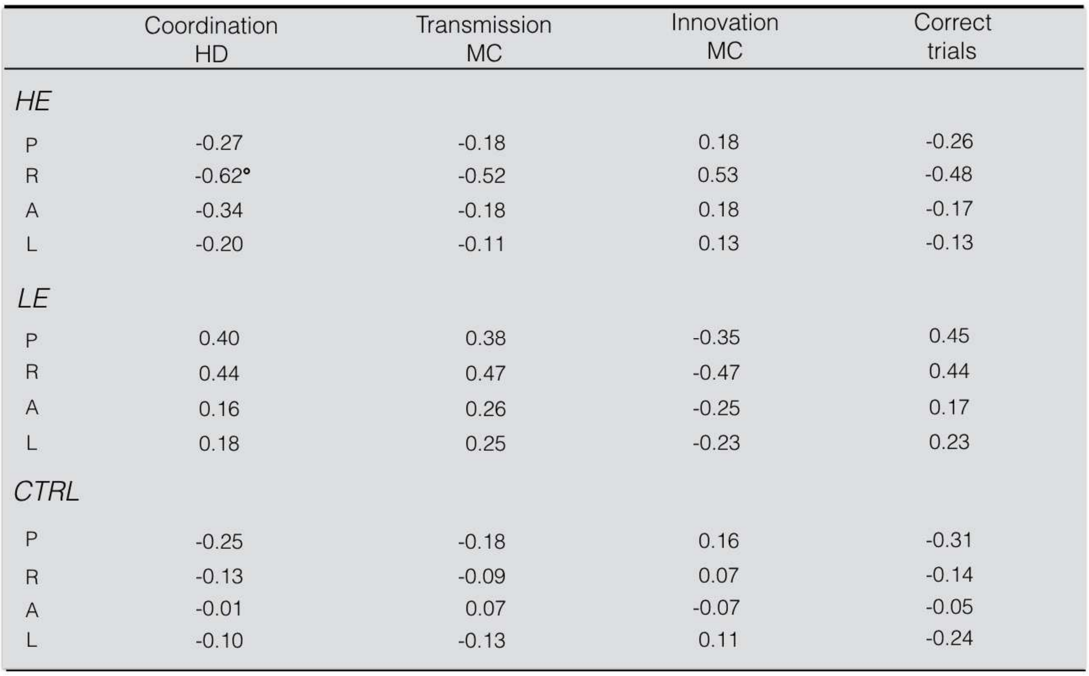
	Coordination HD	Transmission MC	Innovation MC	Correct trials
HE				
P	-0.27	-0.35	-0.42	-0.28
R	-0.30	-0.32	-0.37	-0.26
A	-0.68*	-0.60°	-0.62°	-0.64*
L	-0.58°	-0.51	-0.55	-0.50
LE				
P	-0.40	-0.40	-0.39	-0.31
R	-0.12	-0.12	-0.12	-0.05
A	0.20	0.02	0.01	0.27
L	-0.06	-0.003	0.01	-0.04
CTRL				
P	0.38	0.18	-0.17	0.51
R	0.31	0.15	-0.15	0.46
A	0.15	0.16	-0.13	0.22
L	0.23	0.10	-0.08	0.34

° $p \leq 0.01$
* $p \leq 0.004$

Supplementary Table 1. Pearson product-moment correlations (*r*) between MMN ***peak latencies*** at 120-240[3] ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***behavioral measures*** (coordination, transmission, innovation, accuracy). Coefficients are shown for each scalp quadrant (where P is posterior, R is right, A is anterior, and L is left). P-values are Bonferroni corrected at □=0.05/12=0.004, where 12 is the number of tests involving each independent variable.

---

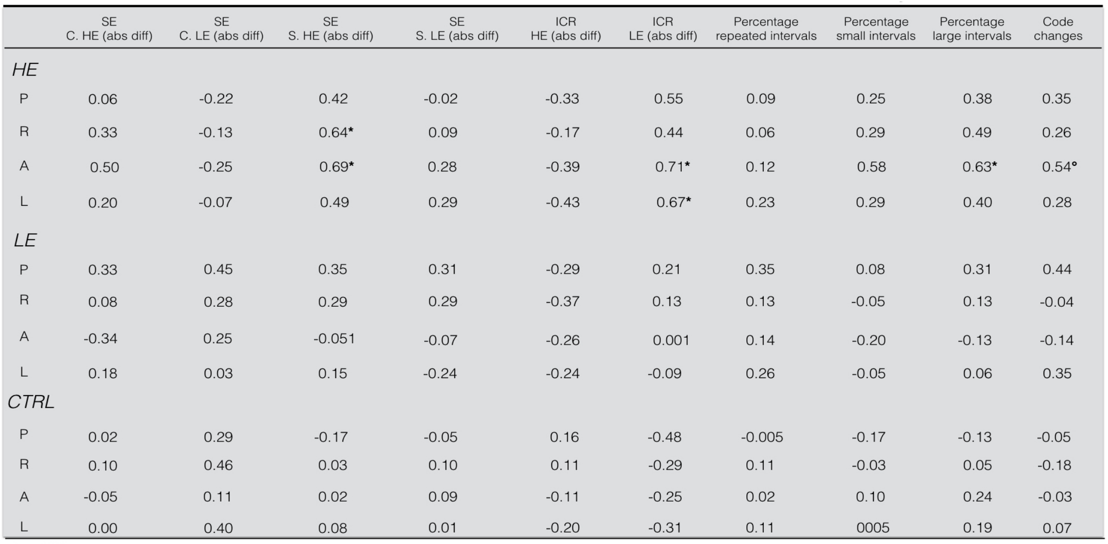
[3] In all cases, the MMN peak fell within the 120-240 ms window. See main article, Fig. 5.



	Coordination HD	Transmission MC	Innovation MC	Correct trials
HE				
P	-0.27	-0.18	0.18	-0.26
R	-0.62°	-0.52	0.53	-0.48
A	-0.34	-0.18	0.18	-0.17
L	-0.20	-0.11	0.13	-0.13
LE				
P	0.40	0.38	-0.35	0.45
R	0.44	0.47	-0.47	0.44
A	0.16	0.26	-0.25	0.17
L	0.18	0.25	-0.23	0.23
CTRL				
P	-0.25	-0.18	0.16	-0.31
R	-0.13	-0.09	0.07	-0.14
A	-0.01	0.07	-0.07	-0.05
L	-0.10	-0.13	0.11	-0.24

° $p \leq 0.01$
* $p \leq 0.004$

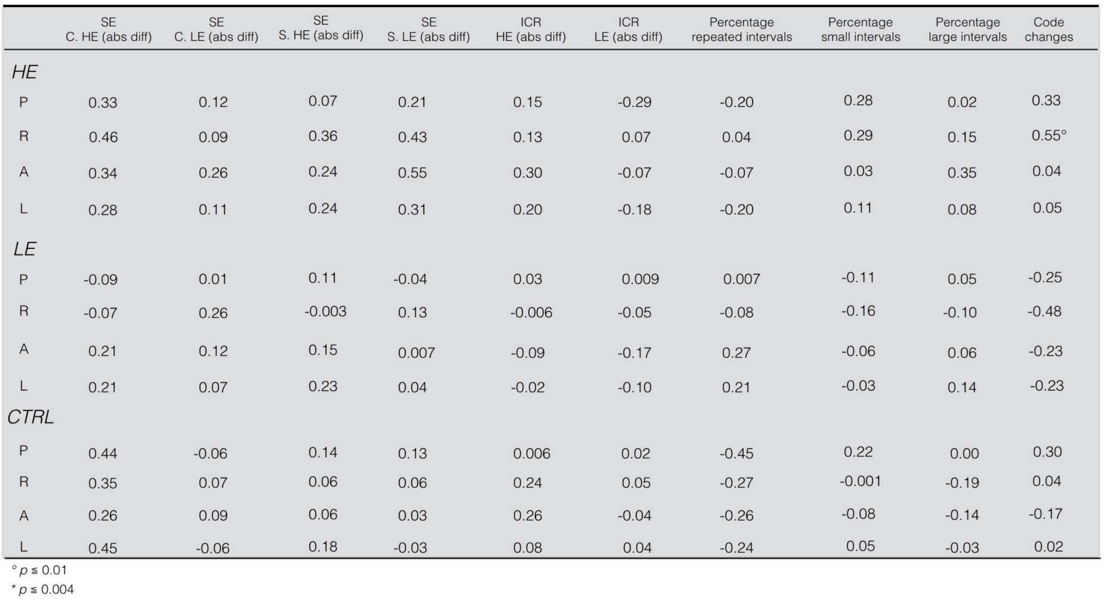
Supplementary Table 2. Pearson product-moment correlations (*r*) between MMN ***amplitudes*** at 120-240 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***behavioral measures*** (coordination, transmission, innovation, accuracy). Coefficients are shown for each scalp quadrant (P is posterior, R is right, A is anterior, L is left). P-values are Bonferroni corrected at □=0.05/12=0.004.



	SE C. HE (abs diff)	SE C. LE (abs diff)	SE S. HE (abs diff)	SE S. LE (abs diff)	ICR HE (abs diff)	ICR LE (abs diff)	Percentage repeated intervals	Percentage small intervals	Percentage large intervals	Code changes
HE										
P	0.06	-0.22	0.42	-0.02	-0.33	0.55	0.09	0.25	0.38	0.35
R	0.33	-0.13	0.64*	0.09	-0.17	0.44	0.06	0.29	0.49	0.26
A	0.50	-0.25	0.69*	0.28	-0.39	0.71*	0.12	0.58	0.63*	0.54°
L	0.20	-0.07	0.49	0.29	-0.43	0.67*	0.23	0.29	0.40	0.28
LE										
P	0.33	0.45	0.35	0.31	-0.29	0.21	0.35	0.08	0.31	0.44
R	0.08	0.28	0.29	0.29	-0.37	0.13	0.13	-0.05	0.13	-0.04
A	-0.34	0.25	-0.051	-0.07	-0.26	0.001	0.14	-0.20	-0.13	-0.14
L	0.18	0.03	0.15	-0.24	-0.24	-0.09	0.26	-0.05	0.06	0.35
CTRL										
P	0.02	0.29	-0.17	-0.05	0.16	-0.48	-0.005	-0.17	-0.13	-0.05
R	0.10	0.46	0.03	0.10	0.11	-0.29	0.11	-0.03	0.05	-0.18
A	-0.05	0.11	0.02	0.09	-0.11	-0.25	0.02	0.10	0.24	-0.03
L	0.00	0.40	0.08	0.01	-0.20	-0.31	0.11	0005	0.19	0.07

° $p \leq 0.01$
* $p \leq 0.004$

Supplementary Table 3. Pearson product-moment correlations (r) between MMN ***peak latencies*** at 120-240 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***structural measures*** (SE is Shannon Entropy, ICR is interval of compression ratio; C. is contour, S. is tone sequence). Coefficients are shown for each scalp quadrant (P is posterior, R is right, A is anterior, L is left). P-values are Bonferroni corrected at α=0.05/12=0.004.

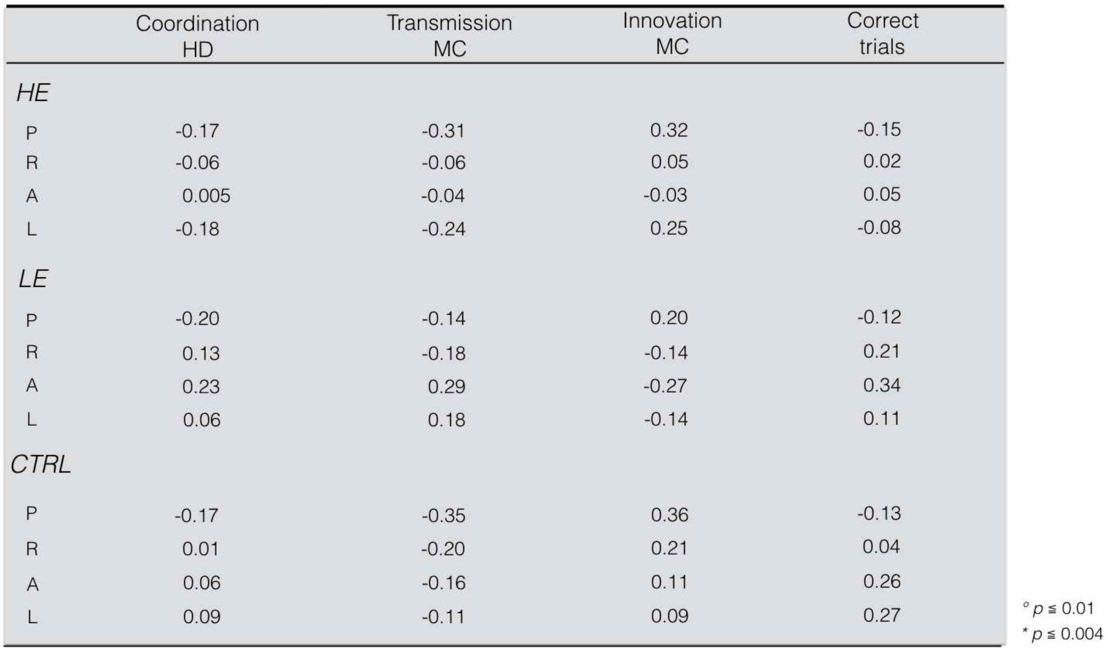


	SE C. HE (abs diff)	SE C. LE (abs diff)	SE S. HE (abs diff)	SE S. LE (abs diff)	ICR HE (abs diff)	ICR LE (abs diff)	Percentage repeated intervals	Percentage small intervals	Percentage large intervals	Code changes
HE										
P	0.33	0.12	0.07	0.21	0.15	-0.29	-0.20	0.28	0.02	0.33
R	0.46	0.09	0.36	0.43	0.13	0.07	0.04	0.29	0.15	0.55°
A	0.34	0.26	0.24	0.55	0.30	-0.07	-0.07	0.03	0.35	0.04
L	0.28	0.11	0.24	0.31	0.20	-0.18	-0.20	0.11	0.08	0.05
LE										
P	-0.09	0.01	0.11	-0.04	0.03	0.009	0.007	-0.11	0.05	-0.25
R	-0.07	0.26	-0.003	0.13	-0.006	-0.05	-0.08	-0.16	-0.10	-0.48
A	0.21	0.12	0.15	0.007	-0.09	-0.17	0.27	-0.06	0.06	-0.23
L	0.21	0.07	0.23	0.04	-0.02	-0.10	0.21	-0.03	0.14	-0.23
CTRL										
P	0.44	-0.06	0.14	0.13	0.006	0.02	-0.45	0.22	0.00	0.30
R	0.35	0.07	0.06	0.06	0.24	0.05	-0.27	-0.001	-0.19	0.04
A	0.26	0.09	0.06	0.03	0.26	-0.04	-0.26	-0.08	-0.14	-0.17
L	0.45	-0.06	0.18	-0.03	0.08	0.04	-0.24	0.05	-0.03	0.02

° $p \leq 0.01$

* $p \leq 0.004$

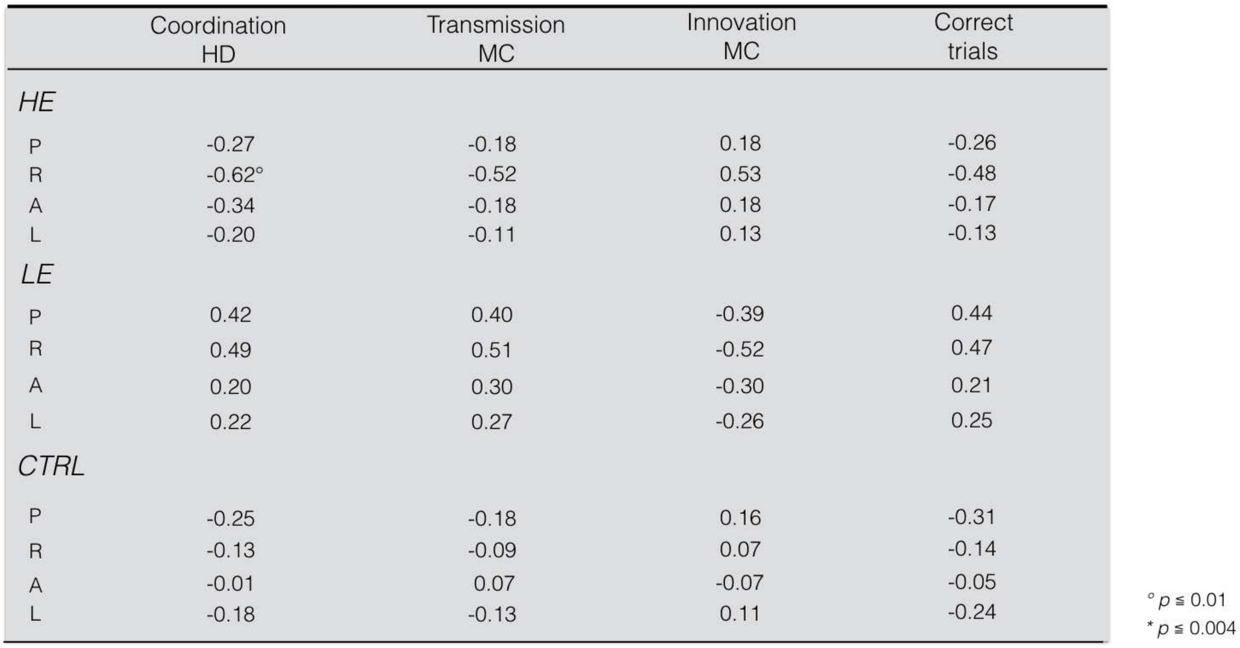
Supplementary Table 4. Pearson product-moment correlations (r) between MMN ***amplitudes*** at 120-240 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***structural measures*** (SE is Shannon Entropy, ICR is interval of compression ratio; C. is contour, S. is tone sequence). Coefficients are shown for each scalp quadrant (P is posterior, R is right, A is anterior, L is left). P-values are Bonferroni corrected at □=0.05/12=0.004.



	Coordination HD	Transmission MC	Innovation MC	Correct trials
HE				
P	-0.17	-0.31	0.32	-0.15
R	-0.06	-0.06	0.05	0.02
A	0.005	-0.04	-0.03	0.05
L	-0.18	-0.24	0.25	-0.08
LE				
P	-0.20	-0.14	0.20	-0.12
R	0.13	-0.18	-0.14	0.21
A	0.23	0.29	-0.27	0.34
L	0.06	0.18	-0.14	0.11
CTRL				
P	-0.17	-0.35	0.36	-0.13
R	0.01	-0.20	0.21	0.04
A	0.06	-0.16	0.11	0.26
L	0.09	-0.11	0.09	0.27

° $p \leq 0.01$
* $p \leq 0.004$

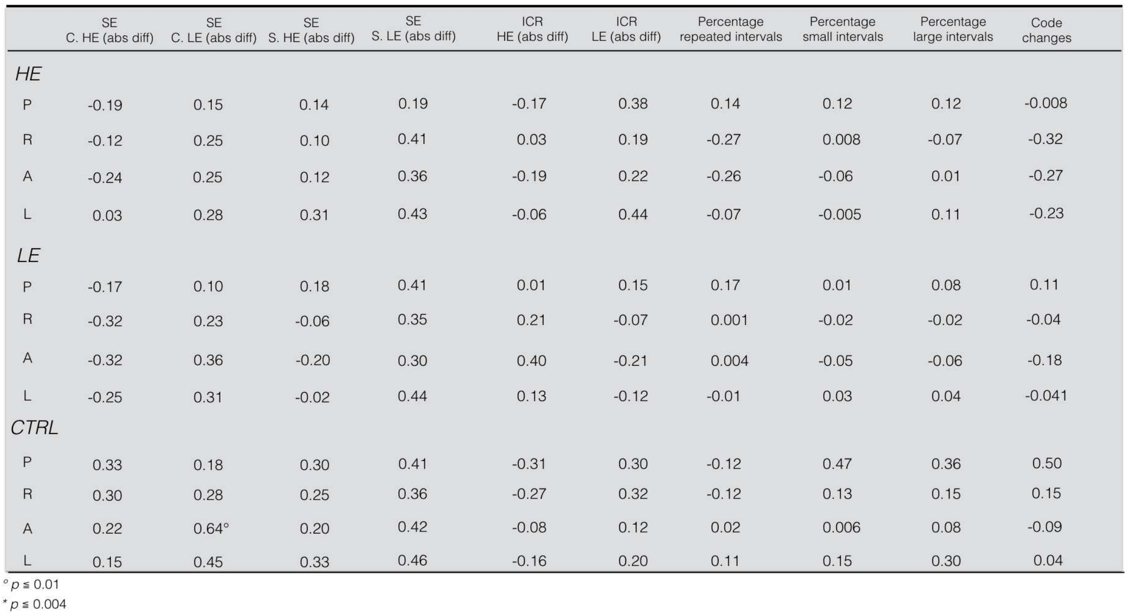
Supplementary Table 5. Pearson product-moment correlations (*r*) between ERP ***peak latencies*** at 240-360 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***behavioral measures*** (coordination, transmission, innovation, accuracy). Coefficients are shown for each scalp quadrant (P is posterior, R is right, A is anterior, L is left). P-values are Bonferroni corrected at □=0.05/12=0.004.



	Coordination HD	Transmission MC	Innovation MC	Correct trials
HE				
P	-0.27	-0.18	0.18	-0.26
R	-0.62°	-0.52	0.53	-0.48
A	-0.34	-0.18	0.18	-0.17
L	-0.20	-0.11	0.13	-0.13
LE				
P	0.42	0.40	-0.39	0.44
R	0.49	0.51	-0.52	0.47
A	0.20	0.30	-0.30	0.21
L	0.22	0.27	-0.26	0.25
CTRL				
P	-0.25	-0.18	0.16	-0.31
R	-0.13	-0.09	0.07	-0.14
A	-0.01	0.07	-0.07	-0.05
L	-0.18	-0.13	0.11	-0.24

° $p \leq 0.01$
* $p \leq 0.004$

Supplementary Table 6. Pearson product-moment correlations (*r*) between ERP ***amplitudes*** at 240-360 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***behavioral measures*** (coordination, transmission, innovation, accuracy). Coefficients are shown for each scalp quadrant (P is posterior, R is right, A is anterior, L is left). P-values are Bonferroni corrected at □=0.05/12=0.004.

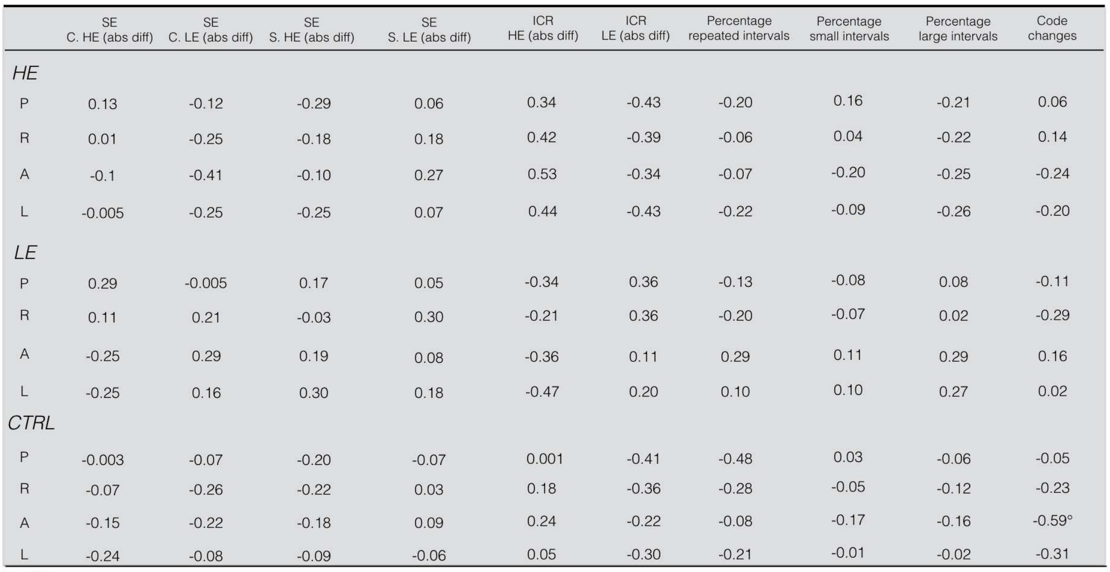


	SE C. HE (abs diff)	SE C. LE (abs diff)	SE S. HE (abs diff)	SE S. LE (abs diff)	ICR HE (abs diff)	ICR LE (abs diff)	Percentage repeated intervals	Percentage small intervals	Percentage large intervals	Code changes
HE										
P	-0.19	0.15	0.14	0.19	-0.17	0.38	0.14	0.12	0.12	-0.008
R	-0.12	0.25	0.10	0.41	0.03	0.19	-0.27	0.008	-0.07	-0.32
A	-0.24	0.25	0.12	0.36	-0.19	0.22	-0.26	-0.06	0.01	-0.27
L	0.03	0.28	0.31	0.43	-0.06	0.44	-0.07	-0.005	0.11	-0.23
LE										
P	-0.17	0.10	0.18	0.41	0.01	0.15	0.17	0.01	0.08	0.11
R	-0.32	0.23	-0.06	0.35	0.21	-0.07	0.001	-0.02	-0.02	-0.04
A	-0.32	0.36	-0.20	0.30	0.40	-0.21	0.004	-0.05	-0.06	-0.18
L	-0.25	0.31	-0.02	0.44	0.13	-0.12	-0.01	0.03	0.04	-0.041
CTRL										
P	0.33	0.18	0.30	0.41	-0.31	0.30	-0.12	0.47	0.36	0.50
R	0.30	0.28	0.25	0.36	-0.27	0.32	-0.12	0.13	0.15	0.15
A	0.22	0.64°	0.20	0.42	-0.08	0.12	0.02	0.006	0.08	-0.09
L	0.15	0.45	0.33	0.46	-0.16	0.20	0.11	0.15	0.30	0.04

° $p \leq 0.01$

* $p \leq 0.004$

Supplementary Table 7. Pearson product-moment correlations (r) between ERP ***peak latencies*** at 240-360 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***structural measures*** (SE is Shannon Entropy, ICR is interval of compression ratio; C. is contour, S. is tone sequence). Coefficients are shown for each scalp quadrant (P is posterior, R is right, A is anterior, L is left). P-values are Bonferroni corrected at □=0.05/12=0.004.



	SE C. HE (abs diff)	SE C. LE (abs diff)	SE S. HE (abs diff)	SE S. LE (abs diff)	ICR HE (abs diff)	ICR LE (abs diff)	Percentage repeated intervals	Percentage small intervals	Percentage large intervals	Code changes
HE										
P	0.13	-0.12	-0.29	0.06	0.34	-0.43	-0.20	0.16	-0.21	0.06
R	0.01	-0.25	-0.18	0.18	0.42	-0.39	-0.06	0.04	-0.22	0.14
A	-0.1	-0.41	-0.10	0.27	0.53	-0.34	-0.07	-0.20	-0.25	-0.24
L	-0.005	-0.25	-0.25	0.07	0.44	-0.43	-0.22	-0.09	-0.26	-0.20
LE										
P	0.29	-0.005	0.17	0.05	-0.34	0.36	-0.13	-0.08	0.08	-0.11
R	0.11	0.21	-0.03	0.30	-0.21	0.36	-0.20	-0.07	0.02	-0.29
A	-0.25	0.29	0.19	0.08	-0.36	0.11	0.29	0.11	0.29	0.16
L	-0.25	0.16	0.30	0.18	-0.47	0.20	0.10	0.10	0.27	0.02
CTRL										
P	-0.003	-0.07	-0.20	-0.07	0.001	-0.41	-0.48	0.03	-0.06	-0.05
R	-0.07	-0.26	-0.22	0.03	0.18	-0.36	-0.28	-0.05	-0.12	-0.23
A	-0.15	-0.22	-0.18	0.09	0.24	-0.22	-0.08	-0.17	-0.16	-0.59°
L	-0.24	-0.08	-0.09	-0.06	0.05	-0.30	-0.21	-0.01	-0.02	-0.31

° $p \leq 0.01$
* $p \leq 0.004$

Supplementary Table 8. Pearson product-moment correlations (r) between MMN ***amplitudes*** at 240-360 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***structural measures*** (SE is Shannon Entropy, ICR is interval of compression ratio; C. is contour, S. is tone sequence). Coefficients are shown for each scalp quadrant (P is posterior, R is right, A is anterior, L is left). P-values are Bonferroni corrected at □=0.05/12=0.004.

Author contributions

Conceived and designed the experiments: ML and GB. Performed the experiments: ML. Analyzed the data: ML. Wrote the paper: ML and GB.

Chapter 4

Neural predictors of temporal regularization

Lumaca, M., Haumann, N.T., Brattico, E., Vuust, P., Baggio, G., (in progress). Neural predictors of learning, transmission, and temporal regularization in a laboratory model of cultural evolution.

Abstract

It has been recently proposed that universal regularities in human symbolic systems may emerge in cultural transmission as adaptations to invariant aspects of brain function. While there is now support from experimental research, no study addressed this hypothesis with neurophysiological methods. How does symbolic structure and symbolic behavior reflect the information processing capacity and constraints of the human brain? Can inter-individual differences in information processing account for aspects of the kind of variation observed in symbolic systems? We addressed these issues in two independent experiments, focusing on musical rhythm. In the first experiment we investigated how regular temporal patterns emerge and change in the course of iterated signaling games, where the receiver in one game becomes the sender in the next game. In the second study, we combined signaling games and electrophysiology in two successive days. On day 1, we recorded MMN responses by participants listening to standard and deviant 5-equitone temporal sequences. On day 2, participants learned rhythmical systems from a trained confederate of the experimenters in one signaling game, and were asked to transmit what they learned by memory in a successive game. The MMN latency from day 1 predicted learning, transmission and temporal regularization of the artificial symbolic system on day 2. This work supports the view that cognitive constraints impose regularity on the symbolic system transmitted, and links regularity of cultural information with constraints on the functional organization of the human brain. Perhaps more interesting, it suggests that small individual neural biases in the reorganization of cultural material can be amplified by cultural transmission to produce universal patterns of human culture.

4.1. Introduction

Music shows a striking diversity across and within cultures, with a multiplicity of variants for each given musical trait (Nettl, 1983; Rzeszutek, et al., 2012). However, it also shows deep similarities, with few of these traits (and variants) widely distributed across world cultures. Examples are the small pitch distance between adjacent tones (named ‘proximity’), and the use of near-equally spaced beats, or ‘isochronicity’ (Savage et al., 2015). How might we reconcile this ostensible contradiction? One recent proposal (Dehaene & Cohen, 2007) aims at explaining specific universal aspects of symbolic domains, such as reading and arithmetic, as a reflection of universal properties of neural function. Cultural systems, as they are introduced in human evolutionary history, ‘invade’ pre-existing neural circuits originally evolved for a different, but related, purpose. As the neural circuits accommodate the novel cultural function, they reorganize themselves while retaining their original computational capacity and constraints, exerting a powerful influence on the way the new function is represented and recalled. In another influential proposal (Christiansen & Chater, 2008; Deacon, 1997), symbolic systems, in particular language, reflect properties of brain structure and function as a result of adaptation, in the course of successive generations, to the selective pressures imposed by human information processing constraints. Experimental research on language, reading and arithmetics, supports now these hypotheses, with remarkable examples of brain reorganization found during cultural acquisition (Hannagan, et al., 2015; for a review see Dehaene et al., 2015), or the emergence of structure observed in culturally transmitted artificial languages (Kirby et al., 2008).

These proposals begin to explain central invariant aspects in human culture, or statistical universals, but they leave the problem of diversity or variation untouched. How cultural variants originate at first in a population? And how the proportion of these variants changes to produce universal aspects of human culture? Our hypothesis is that differences in brain structure and function across individuals can manifest in differences in cultural behavior, which may spread in the population via cultural transmission mechanisms, enabling universal patterns found in cultural symbolic systems. This view has recently found some support by cultural evolution research. In their work, Dediu & Ladd (2007) show that population frequencies of two human genes, *Microcephalin* and *ASPM*, are reliably associated to the presence (or not) of linguistic tones in that population. The author’s proposal is that allelic variants of the two genes may determine small biases at individual level in the processing and acquisition of linguistic tones. This is not enough to manifest their phenotypic effects, since tonal and non-tonal language can be acquired by people with or without these genes. These variants need to be amplified by cultural transmission across generations (Kirby et al., 2008) to produce large-scale population patterns, here the presence (or absence) of a tonal language. In another study, Lumaca and Baggio (2016), showed that individual differences in neural information processing, revealed by an index of auditory processing efficiency, the mismatch-negativity (or MMN) (Naatanen, 1978), predicted to what extent individuals learned, transmitted and reorganized a simple symbolic system with signals (artificial tone sequences) associated with meanings. These two studies (Dediu & Ladd, 2007; Lumaca & Baggio, 2016) suggest a possible connection between neural variability and universal properties of cultural systems.

Here, we address this hypothesis in the laboratory, with a focus on musical rhythms: (1) we anchor individual neural variability in information processing with small cognitive biases in the acquisition, transmission and reorganization of complex temporal patterns; and (2) we examine whether cultural transmission can amplify these biases, leading to the emergence of an universal property of musical rhythms, namely isochronicity (Fitch, 2016). We did so in two independent studies. We first tested whether isochrony evolves over generations, i.e., whether signals tend to become more regular rhythmically, possibly to facilitate retention, recall and transmission in the diffusion chains. Next, we examined whether individual variation in neural information processing, revealed by ERPs, predicts variation in cultural behavior: specifically, how well participants learn the signal system, how faithfully they re-transmit it, and what kind of structural modifications they introduce in the process.

In the first study, we use *multi-generational signaling games* (MGSGs) (Lewis, 1969; Skyrms, 2010; Moreno & Baggio, 2015; Nowak and Baggio 2016), that is, signaling games played iteratively in diffusion chains (Esper, 1925; Bartlett, 1932), where one game, played by a sender and a receiver, corresponds to a generation. Players must converge on a mapping of signals (equitone temporal patterns) to meanings (here, simple and compound emotions), or *code*, via iterated signaling rounds (Fig. 1). The receiver in each game becomes the sender in the next game, which he will play with a new participant in the diffusion chain (a receiver). This process is repeated until the end of a chain (9 generations). Seeding is used to initialize the chains with controlled material (Whiten et al., 2007; Flynn & Whiten, 2008). We measured code changes across generations in timing regularity (*Shannon Entropy*) and fidelity of transmission (*transmission* and *innovation*). We assessed the direction of information flow in chains using two indexes: *role asymmetry* and *coordination* (Moreno & Baggio, 2015) (see Methods). Our aim was to determine whether tone sequences, denoting specific meanings, would be gradually regularized by successive generations.

In the second study, we use two approaches over two successive days (Lumaca and Baggio 2016): (i) a reduced version of MGSGs and (ii) an event-related potential (ERP) experiment aimed at identifying a neural marker of processing efficiency for structured tone sequences. In the reduced version of MGSGs each individual played two signaling games. In the first game, the subject played as the receiver with a confederate of the experimenters as the sender, and was asked to learn a *seed code*, consisting of given associations of tone sequences and meanings (as in Study 1). In the second game, the participant is asked to retransmit that code to the receiver. We first measured behavioral performance: individual learning scores (*coordination*), transmission fidelity (*transmission* and *innovation*), and accuracy in the transmission of the code over trials (*consistency*). We computed changes in time regularity (*entropy*) between the seed and the reproduced code. Next, on the second day, as a neural marker of processing capacity, we used the mismatch negativity (MMN). The MMN is a negative ERP component peaking between 100 and 220 ms from the onset of a *deviant* (i.e., infrequent) event in a train of *standard* stimuli. The MMN has been often used in the past as neural marker of auditory pattern processing efficiency (Näätänen et al., 1993; Tervaniemi et al., 2001; Alain et al., 1999; Gottselig et al., 2004; Lumaca & Baggio, 2016). We tested whether individual differences in auditory processing, as revealed by MMN peak latencies and amplitudes recorded on Day 1, predict behavioral performance and regularization of tone sequences in signaling games on Day 2. The main contribution of this study is to introduce, albeit with a simplified model, a first experimental link between individual differences in the functional organization of sensory systems and universal properties of musical structure.

4.2. Methods

4.2.1. Experiment 1

4.2.1.1. Participants

Sixty-five volunteers (1 confederate) participated in Experiment 1 (33 female, mean age 24.94). All had normal hearing, with no formal musical training. Upon arrival to the lab, participants were informed that they would play a game at a computer terminal with a partner, as described by written instructions (Suppl. Inf. 1.1). Participants were organized in 8 transmission chains with 9 generations each (the confederate played as first sender in all chains).

4.2.1.2. Stimuli

The *states* in signaling games were 5 emotions: 3 *basic* (joy, peace, sadness), 2 *compound* (peace+joy, peace+sadness), shown as facial expressions. Compound emotions were produced first by dividing up basic emotion faces in two regions: *upper* and *lower face* (Ekman & Friesen, 2003), then by combining the upper face features of peace with the lower face features of joy (peace+joy) and sadness (peace+sadness). In compositionality analyses, the upper face was coded as 2 variants: *open* or *closed* eyes;

the lower face was coded as 3 variants: mouth corners *up, straight* or *down*. Prior to the start of the games, participants were trained to associate a facial expression to a simple or compound emotion, in 3 blocks of increasing complexity. In the first and second blocks, respectively, facial expressions for simple and compound emotions were presented on the screen to the participant. The participant was asked to learn the mapping between facial expressions and emotions, selecting one among the orthographical options (numerical digit 1 to 5) presented on the screen. A feedback followed, showing to the participant the correct answer. After 4 correct responses participants moved to the test, where simple and compound emotions were randomly presented. This time, no feedback followed. The *signals* denoting the emotions were 4-note rhythms of 2 sec duration, each denoting a different state. Sounds were drum timbres of 100-ms duration (5-ms fade in-out) to be produced by participants using one predefined key on the numeric keypad. To aid perception of correct meter, the signal was reproduced twice and delivered through stereo headphones at 80 dB.

4.2.1.3. Procedure

Signaling games were played in an experimental room using two computer terminals facing each other. Sender and receiver were provided with a computer keyboard and stereo headphones. Roles were maintained fixed during each game. At the end of a game, the receiver (generation *n*) became the sender and played with a new receiver (generation *n+1*). The sender was now instructed to transmit as faithfully as possible the code learned in the previous game. Thus, a diffusion chain of 9 generations (8 games) was constructed. Participants were never allowed to communicate verbally or otherwise besides the signaling game itself. The experimenter was always present in the testing room. Fig. 1 shows the structure of a trial. The sender is privately shown a facial expression (2 sec), and sends a rhythmic signal denoting that emotion to the receiver, composing a 4-note sequence (with the exception of the first sender of a chain, who didn't need to compose the signals by himself; see below). Unheard by the receiver, the sender could try a combination of beats at will, listening to the final beat sequence played twice (8 beats; 4 sec) via headphones. Producing signals of the correct duration (max 2 sec) was required to proceed in the game; if not, the signal was automatically rejected, and had to be composed anew. The signal was delivered to the receiver, who listened to it via headphones (repeated twice: 4 sec). The receiver responded by selecting among the 5 facial expressions presented on the screen the one the sender might have seen. The order of facial expressions was randomized across trials. Feedback (2 sec) was then shown simultaneously to the players, showing the face the sender had seen and the one that the receiver had chosen. A trial is correct if the two expressions match. The game ended at 70 trials. Each facial expression was randomly presented 14 times in each session, twice every 10 trials and never in sequence.

We designed 4 signal sets to initialize the 8 transmission chains (or *seeding material*; see Suppl. Tab. 1.3 and Suppl. Fig. 1.4.1): 2 sets (chains 1-4) with *high*-syncopated stimuli (HS; syncopation, M=2; metric complexity, M=2.95, SEM=±0.26), 2 sets (chains 5-8) with *low*-syncopated stimuli (LS; syncopation, M=1; metric complexity, M=1.96, SEM=±0.32). The following parameters were matched across signal sets: regularity (Shannon entropy=1.5, range 0:2), regularity on rhythmic contour (Entropy = 0.91, range 0:1.58) and compression on rhythmic contour (within-set rhythm similarity, M=0.66, range 0:1). See below for information on these measures. In each set of signals, the stimuli varied in the type of metrical categories: duple (16 subdivisions per metrical cycle; 2 rhythms), triple (12 subdivisions; 2 rhythms), irregular meters (15 subdivisions; 1 rhythm). Tempo was always 120 beat per minutes (bpm). A confederate (*generation 1*; first sender in each chain) was instructed to deliver the rhythmic stimuli to the partner (receiver; *generation 2*). For each trial, the confederate was presented on the screen with a facial expression, and, below the picture, a digit corresponding to the key to press on the computer keyboard (1 to 5). Each digit was mapped to one of the 5 four-note rhythm patterns, i.e., the one denoting the expression observed at the beginning of the trial.

4.2.1.4. Data analysis

We aimed to describe the evolution of rhythmic regularities in the course of transmission within diffusion chains. The main dependent variable was *regularity* (measured by entropy) in linear mixed-effects regression models (Winter & Wieling, 2016) in R Studio (R Studio Team, 2015) and *lme4* (Bates et al., 2012). We modeled dependent variables (y) as a function of Generation and Syncopation (fixed effects) with random intercepts (by-chain variation in y) and random slopes for generation (by-chain variation in the slope of generation). No differences were observed between models with (i.e., the full model) and without Syncopation, which was removed from the final analysis ($\chi^2(1)=0.31$, p=0.57). The opposite pattern was instead observed excluding Generation. Specifically, only the likelihood ratio test between the full model and the reduced model without Generation resulted significant ($\chi^2(1)=9.55$, p=.001). Generation is the factor that most explains the data given the model. This justifies our use of a model where Generation is the only fixed effect factor. There were no deviations from homoscedasticity or normality. We tested changes over time in regularity by using likelihood ratio tests for the fixed effect included in the model, against a null model (fixed effect excluded). Non-parametric Wilcoxon signed-rank tests (Siegel & Castellan, 1988) were used to compare data points between generations.

In addition, a baseline measure of change was included, reflecting how codes would evolve by chance (dotted line in the figures). To this purpose the original dataset of the interonset intervals was randomly shuffled (n=1000 times): a one-sample Wilcoxon test was performed between codes produced in the last generation and the median baseline value. To test whether signaling games can capture the cultural transmission process, independent of the initial levels of syncopation, we measured *role asymmetry*, *coordination, transmission* and *innovation* (Moreno & Baggio, 2015; Nowak & Baggio, 2016), pooling the two conditions.

First, to account for temporal imprecisions introduced by players, and to facilitate a comparison, all signals produced were linearly *normalized* to match the original tempo (120 bpm) and *quantized* on the best fitting metrical grid. Normalized rhythms (N) were computed by multiplying the pattern of inter-onset intervals (IOIs) of the played rhythm with a tempo factor. This factor was calculated as the ratio between the time of the last onset in the played rhythm (P) and the time of the last onset in the rhythm of the seeding material (S) denoting the same emotion: N = IOIs * (S/P). A quantized signal was then produced by rounding up (or down) the normalized played onsets to the nearest onset on the best fitting metrical grid (binary, ternary or quinary), i.e., the one that minimized the sum of squared distance with the normalized played onsets.

Changes in the *regularity* of signals were measured using Shannon Entropy (Shannon, 1949) on time intervals. Lower scores are associated to isochronous signals. Other measures (*syncopation*, *metric complexity*, *compression* and *compositionality*) are described in Suppl. Inf. 1.2**.**

We measured the *similarity* between pairs of rhythms using the inverse of the normalized Hamming distance (Hamming, 1986). The normalized Hamming distance is defined as the number of substitutions required to match two strings of equal length, divided the number of elements in the strings. In this study, strings were 4 inter-onset intervals and relative contour transformations: the series of *ups* (larger intervals), *constants*, and *downs* (shorter intervals), independent of their exact size *(rhythmic contour similarity*). The inter-onset interval is known as the temporal parameter encoded in tone sequences (Schulze, 1989; Vos, Mates, & van Kruysbergen, 1995), while the relative contour is the form of temporal organization most recognized (Dowling & Harwood, 1986).

Coordination indicates the fraction of the code shared by the two players at the end of the session. It was measured as the mean similarity of rhythms denoting corresponding states and used with major frequency in the second half of a game by sender and receiver (i.e., when responses by players are more consistent and show less random errors). Similarly, we computed the coordination index for the relative rhythmic contours. Values close to 1 denote a shared signaling system.

One-sample Wilcoxon were used to test the null hypothesis that coordination is not significantly different from 0.

Then, we measured *role asymmetry* (A) between senders and receivers (Selten & Warglien, 2007). This measure indicates the differential effort carried out by the two players to achieve coordination on the signal-to-state mappings. It was measured as the difference in the number of code changes introduced by sender (S) minus the ones of the receiver (R), divided by the total number of code changes: A=(S-R)/(S+R). With negative asymmetry (the receiver changes his code more often than the sender), information flows from sender to receiver. Moreno and Baggio (2015) show that this condition is achieved if player roles are fixed during the signaling games. In this condition, signaling games become a viable model of code transmission (Nowak & Baggio, 2016). One-sample Wilcoxon tests were used to test the null hypothesis that mean role asymmetry is not significantly different from zero.

Transmission indicates how faithful the transmission of the signaling system is between two successive generations. It was measured as the mean similarity of rhythms (or their contour transforms) denoting corresponding states, produced with greater frequency in the second half of the game by senders of generation n and generation $n+1$ (between-player measure). *Innovation* indicates to what extent a code changed between two successive games by the same players (within-player measure). It is measured as the mean normalized Hamming distance between codes learned by the receiver in generation n and codes reproduced by the same person playing as sender in generation n+1. Only when coordination is equal to 1, innovation and transmission mirror each other. Changes in transmission and innovation over generations were tested using a linear mixed-effects regression model, with mean transmission and innovation values as dependent variables respectively.

4.2.2. Experiment 2

4.2.2.1. Participants

Seventeen volunteers were recruited for the EEG study (11 females; mean age = 25.9 y, SD = 5.6; 5 additional participants were excluded due to excessive EEG artifacts). To avoid effects of musical training on auditory processing (Vuust et al. 2009), musically naive subjects were enrolled. All participants signed a written information consent sheet (Suppl. Inf. 2.1).

4.2.2.2. Design

Each subject participated in 2 experimental sessions on 2 successive days (24h apart). On day 1 the EEG session took place. The EEG was recorded in three blocks with rhythmic patterns at 3 different levels of regularity (see section 4.2.2.3.1) delivered by loudspeakers (80 dB). In the meanwhile, the subject was watching a subtitled silent movie centered on the screen (8x11 frame). Subjects were specifically instructed to ignore the sounds coming from the loudspeaker and pay attention to the movie. They were also asked to minimize movements, including eye movements and blinks. On day 2, the subject played 2 signaling games with a confederate, the first as *receiver*, the second as *sender.* In game 1 the participant was instructed to learn a signaling system from a confederate. In game 2, the participant (now sender) was instructed to re-transmit the signaling system to the confederate as they best remembered it (now the receiver). Sender and receiver roles were fixed during the session. The signaling system consisted of 5 four-note rhythms denoting affective meanings (3 simple and 2 compound emotions). Two rhythms were high-entropic (H=2 bits), 2 low-entropic (H=1 bits) and 1 control (H=0 bits; isochronous). Metric subdivisions (1/16), duration (2 sec), tempo (120 bpm) and syncopation levels (0) were matched across stimuli. Three sets of signals with these parameters were built and counterbalanced across participants.

4.2.2.3. Day 1. EEG

4.2.2.3.1. Stimuli

The stimuli consisted of sine-waves equitone rhythms of 5 beats, presented in 3 recording sessions. The order of sessions or conditions (high-entropy, H=2 bits, low-entropy, H=1 bits, and control, H=0 bits) was counterbalanced across participants. The Shannon Entropy of the stimuli indicates the average minimum information (in bits) necessary to encode the rhythmical sequence, and provides a reliable measure of rhythmical complexity. Low entropic rhythmic sequences can be encoded in fewer chunks of information, and are easier to learn and store in memory. Each session was approximately 15 min long, and consisted of 4 blocks with 100 stimuli each, for a total of 1200 stimuli per session. Fig. 3 shows the 3 types of stimuli used in each session: 1 standard and 2 deviants. In each block of the session, 80% were standard stimuli 1300-ms long with 700-ms of interstimulus interval (ISI) (high- and low-entropic stimuli) or 1600-ms long with 900-ms of ISI (control stimuli). On 10% of the trials, early deviants were presented in which the fourth beat was shifted 300-ms earlier, altering the rhythmic contour (*contour deviant*). On the remaining 10% of trials, the fourth beat was anticipated by 100-ms, producing a deviant that preserved the original rhythmic contour (*timing deviant*). Standards and deviants were presented in pseudorandom order, with the only constraint that two deviants could not occur in a row. The equitone rhythms within each block were randomly transposed at three different fundamental frequencies (315.0 Hz, 397 Hz and 500 Hz).

4.2.2.3.2. EEG recording and processing

EEG data was recorded at a sampling rate of 1000 Hz from an EEG actiCap (Brain Products) with 64 Ag-AgCl electrodes arranged in the international 10-20 system and referenced to FCz. EEG data were offline downsampled to 500 Hz, bandpass filtered between 0.1 and 30 Hz (roll-off = 12 dB/octave) using the Matlab toolboxes ERPlab (Lopez-Calderon and Luck 2014) and EEGlab (Delorme and Makeig 2004). The EEG was re-referenced offline to the average of the mastoid channels and then segmented into 700-ms epochs (-100 to 600 ms relative to the onset of the fourth beat). EEG responses exceeding ±70 μV (peak-to-peak; moving window = 200 ms) in any epoch were considered artifacts and excluded from the average.

4.2.2.3.3. ERP data analysis

The remaining epochs were averaged separately for each condition (standard and deviants), and deviant-standard difference waves were computed for each participant by subtracting standard from deviant ERPs. This procedure isolates the mismatch negativity (MMN) component, and possibly other effects, in ERPs. The MMN is a negative ERP component fronto-centrally distributed and peaking between 100 and 220 ms in response to an auditory mismatch (Näätänen et al., 1978).

First, we tested the presence of an MMN using a four-way repeated measures ANOVA with mean MMN amplitude as dependent variable in each of 3 conditions, and 4 factors: Stimulus Type (standard or deviant), Temporal Window (4 levels), Quadrant (4 levels) and, nested on the latter, Electrode (9*4 levels). The Greenhouse-Geisser correction was applied when appropriate. Wilcoxon signed-rank tests were used in post-hoc analyses. The 4 scalp quadrants were: anterior, central left, central right and posterior (9 channels each). Four 80-ms time windows (0-320) were selected according to the timing characteristics of the MMN and after inspection of the ERP data.

Then, we tested whether the MMN recorded on day 1 could predict learning, transmission and reorganization of signaling systems on day 2. To this purpose, we correlated individual peak latencies or amplitudes of MMN difference waves obtained in each time window/quadrant, with different measures of behavior in signaling games (see below). Based on our previous work (Lumaca & Baggio, 2016) we expected that ERPs with the timing and topographical characteristics of the MMN be consistently predictive of behavioral measures.

4.2.2.4. Day 2. Signaling games

4.2.2.4.1. Stimuli

Signals and states were the same as in Experiment 1. However, in Experiment 2 the signals were matched in metric subdivision (binary; 16 subdivisions) and syncopation (S=0). In each set of signals, stimuli varied in regularity: 2 rhythms were high-entropic (H=2 bits), 2 low-entropic (H=1 bits) and 1 isochronous (H=0 bits). Sounds were produced as sine-waves at a fundamental frequency of 500 Hz. Three set of stimuli with these properties were produced and counterbalanced across participants as starting material in game 1.

4.2.2.4.2. Procedure

The trial design was the same as in Experiment 1. The duration of game 1 was fixed at 70 trials. Game 2 ended when 3 correct trials in a row for each emotion was achieved, with a limit of 50 trials (mean=33.05; SD=5.75).

4.2.2.4.3. Data analysis

We tested whether MMN amplitudes or latencies predicted individual learning, transmission and regularization of signaling systems. The same procedures (*tempo normalization* and *quantization*) and formal measures used in Experiment 1 (*asymmetry*, *transmission*, *innovation* and *regularity*) were applied to the set of signals in Experiment 2. Only here, a more sensitive measure of coordination was computed on the data (Lumaca & Baggio, 2016). Coordination is here defined as the mean contour similarity between the signal used by the confederate for a given emotion and the set of signals used by the participant for the same emotion during the second half of the game 1, where player responses are more consistent. In addition, the *consistency* of a participant in sending signals in game 2 was computed as the mean similarity between contours of rhythms produced for each given emotion. Values close to 1 indicate high consistency: signals to state associations do not change during a game. We computed Pearson-product moment correlations (*r*) between behavior (coordination, transmission, innovation, consistency) and structural changes across games (regularity), and MMN parameters.

4.3. Results

4.3.1. Result Experiment 1

Fig. 2 shows entropy changes across generations. A likelihood ratio test of the reduced model against the null model reveals a significant trend of decreasing entropy ($\chi^2(1)=9.45$, p=.002). This is supported by a significant difference in entropy between the first and the last generations (Wilcoxon signed-rank test, p=0.01). This indicates that the initial rhythmic material was regularized over time. Moreover, a significant difference in entropy was observed between the last generation and median baseline levels (one sample Wilcoxon, p=.005). This suggests that changes in entropy are not random. We did not observe any significant predictive effects of generation or significant differences from baseline values for others structural measures (*syncopation*, *metric complexity*, *compression* and *compositionality*), that remained constant over time (Suppl. Figs. 1.4.2.-1.4.5).

Asymmetry was negative and significantly different from 0 (median=-0.13, n=64, Z=-6.957, p<0.001): receivers adjusted their mappings more often than senders did during coordination. *Coordination* (on rhythmic contours) was also significantly different from 0 (median=0.51,n=64, Z=6.955, p<0.001): most of the code was shared between players at the end of each session. No significant changes over generations were found for contour *transmission* and *innovation*, comparing the reduced and null models (transmission, $\chi^2(1)=0.005$, *p*=0.93; innovation, $\chi^2(1)=0.62$, p=0.42). The high levels of transmission observed in generation 2 (M=0.66, SD=0.11) were maintained throughout the diffusion chain, until the last game (M=0.67, SD=0.19). We observed similar results for the low values of *innovation* (G2, M=0.45, SD=0.20; g8, M=0.35, SD=0.31). The high values of transmission and the low levels of innovation suggest that most of the code was transmitted from one generation to the next, and that only a small fraction of it was changed by participants when they played the game as senders. When we used the same metrics for rhythmic codes with IOIs values, we observed a low coordination value, significantly different from 0 (median=0.24, n=64, Z=6.955, p<0.001), and no significant changes either in transmission or in innovation (transmission, $\chi^2(1)=0.14$, *p*=0.78; innovation, $\chi^2(1)=0.73$, p=0.40).

4.3.2. Result Experiment 2

4.3.2.1. Signaling games

Asymmetry during game 2 was negative and different from 0 (one sample Wilcoxon, median=-0.05, n=17, Z=-3.18, *p*<.001). Coordination was also different from 0 (one sample Wilcoxon, median=0.91, n=17, Z=4.01, *p*<.001). These data indicate that receivers adjusted their mappings more often than senders, and that learning of the codes occurred during game 1.

Transmission of rhythms was high and significantly different from 0 (taken on rhythm contours; one sample Wilcoxon, median=0.7, n=17, Z=3.62, *p*<.001) with scarce innovation (one sample Wilcoxon, median=0.2, n=17, Z=3.34, p=.001). An opposite pattern was observed using the Hamming distance of inter-onset intervals, with low levels of transmission (0.35) and high levels of innovation (0.7). These results highlight the privileged status of rhythmic contour in temporal perception and memory, and at the same time, they suggest that interonset interval sequences are not a reliable measure of rhythmic learning. This pattern of results justifies our use of rhythmic contour in subsequent analysis, where we are trying to relate reliable indexes of rhythmic learning with indexes of neural information processing capabilities.

Finally, we observed a consistent reorganization of the structure of signaling systems by the participants. The entropy of the signals changed significantly between game 1 and 2 (Wilcoxon signed-rank test, Z=3.18, *p*<.001). We observed an increase of entropy for low-entropic rhythms (Wilcoxon signed-rank test, median=0.5, Z=-3.64, *p*<.001), with no significant changes for isochronous stimuli (Wilcoxon signed-rank test, median=0, Z=-1.89, *p*<.001). A different pattern of results was observed for the high-entropic rhythms: entropy significantly decreased between the 2 games (Wilcoxon signed-rank test, median=0.25, Z=-3.496, *p*<.001).

4.3.2.2. MMN

Timing and topographical distribution of the event-related potential elicited by deviants are consistent with the MMN (Figs. 4-5). Results in the ANOVA and post-hoc analyses on interaction effects show that the largest effect was in the central-left quadrant between 160 and 240 ms in all conditions (Table 1).

We further examined whether ERPs recorded on day 1 predicted participant behavior in signaling games on day 2. To this purpose we correlated MMN amplitudes and latencies with signaling measures and changes in the regularity of signals between the two games, as reported above. We focus on the central left quadrant at 160-240 ms from the deviant onset, which showed the stronger effects in the previous analysis.

4.3.2.3. Neural predictors

Fig. 6 shows correlations between the MMN peak latencies (160-240 ms) from the central left quadrant and behavioral (*coordination*, *transmission*, *innovation*, *consistency*) and structural measures (*regularity*) (see also Suppl. Tabs. 2.3.1-2.3.2). The MMN peak latency was negatively correlated with *coordination* (p=.006), *transmission* (*p*<.001) and *consistency* (p=.005), and positively related with *innovation* (p=.007): players with shorter MMN peak latencies tended to learn and re-transmit more efficiently the signaling system than players with longer peak latencies.

Changes in the *regularity* of signaling systems were also predicted by MMN peak latencies. Positive correlations were found with mean absolute changes in the entropy of signaling systems (p=.007) and mean changes of high-entropic signals (p=.005). These data reflect a general regularization of the code, mainly introduced by individuals with longer peak latencies. The same individuals were found to be most responsible for the regularization of highly irregular signals. No correlations were found in other conditions (and quadrants) (Suppl. Tabs. 2.3.1-2.3.2). Moreover, no significant correlations were found using MMN amplitudes (Suppl. Tabs. 2.3.3-2.3.4). We did not observe any significant correlation (peak latency or amplitude) in the 240-320 ms post-MMN window (Suppl. Tabs. 2.3.5-2.3.8).

4.4. Discussion

We examined the link between cultural and neural variability in the temporal domain. We demonstrate that near-periodic rhythms arises in signaling systems when these are transmitted across generation of learners via signaling rounds. Then, we show that learning, transmission and temporal regularization of tone sequences by single individuals relate to individual (pre-attentive) information processing capabilities. This result supports and extends the previous findings by (Lumaca and Baggio 2016) on the temporal dimension. Below we discuss the implication of these findings on the broader context of cultural evolution.

4.4.1. The cultural evolution of isochronicity in the laboratory

Rhythmical structure implies large regularity and predictability of temporal events (Huron, 2006; Levitin, Chordia, & Menon, 2012; Patel, 2010; Vuust et al., 2009). Diffusion chains studies have shown that regularity in cultural symbolic systems, such as language grammar and color terminology (e.g., Kirby et al., 2008; Xu, Dowman, & Griffiths, 2013), and other systems (e.g., Kalish, Griffith, & Lewandowsky, 2007), may reflect universal principles of cognition and perception. In this regard, constraints on temporal processing and encoding, such as memory space, processing time, and perceptual categorization (Drake & Bertrand, 2001) might be responsible for the origins of temporal regularity in music (Merker et al., 2015; Trehub, 2015). Irregular rhythmic patterns are perceived as regular if their time intervals remain within certain tolerance windows. As a result, complex durations ratios (2:3 or 2:5) are perceptually reduced onto simpler interval ratios (1:1, 1:2, or 1:3). This also facilitates human memory, which must retain and processe a lower number of interval categories (three or four) (Fraisse, 1946; Pablos Martin et al., 2007; Povel, 1981). Repeated episodes of perceptual categorization and memory processes, during cultural transmission, and the rhythmic sequences are expected to become more regular over time. The progressive decrease of rhythmic entropy found in the first study supports this hypothesis. In our study, cultural information is progressively transformed or distorted towards rhythmic sets of fewer (1 or 2) durational categories, i.e., more regular rhythmic set. These interval categories are widely distributed across the musical cultures of the world (Fraisse, 1956, 1982). Similar findings were obtained Ravignani et al. (2016) using a different model of cultural transmission, the *iterated learning* (Kirby et al., 2008). In Ravignani et al. (2016), a participant (or experimental generation) was there instructed to learn and reproduce a set of random computer-generated patterns, by using a drumstick and an electronic pad. The output produced was stored on a computer device, and used by another participant (the next generation) as training input. The iteration of this procedure allowed the experimenters to observe changes of the rhythmic patterns over time (i.e., over generations). As a result, these patterns increased in regularities. The authors observed the emergence of a smaller set of duration ratios and regular timing (isochronicity). The conceptual reproducibility of this result supports the view that temporal regularity in music may have arisen by economical processing principles (Chater & Vitányi, 2003), not necessarily specific to music (Cornish et al., 2013). Temporal regularity is a design feature of music, very typical of ensemble and dance around the world. In these contexts, multiple individuals must coordinate their listening and actions to precise events in time, and, because of that, some predictable tempo becomes invaluable. This raised the hypothesis that isochronicity may be a byproduct of pressures for synchronization between different individuals during group performance (Fitch, 2012). Savage et al. (2015) found that traditions with isochronous beat, also tend to have group performances, a universal relationship that supports the idea by Fitch (2012). The results of our study, with the support of Ravignani et al. (2016), add a new ingredient on the list of the putative mechanisms at the origins of this feature. Cognitive limits on cultural transmission, in addition to social pressures for synchronization (Cornish et al., 2013; Cross et al., 2013; Fitch, 2012), might be a driving force behind the emergence of temporal regularity in musical systems.

4.4.2. A neurophysiological index of individual constraints on temporal processing

The auditory MMN reflects the brain's automatic detection of deviations from regular representations, and how accurately these representations are encoded in memory (Alain, Achim, et al., 1999; Alain, Cortese, et al., 1999; Winkler, 2007; Winkler, Denham, et al., 2009). The MMN found here provides further evidence that temporal structure is one of the elements encoded in auditory predictive models (e.g., Ford & Hillyard, 1981; Geiser, Ziegler, Jancke, & Meyer, 2009; Kujala et al., 2001; Port, 1991;

Sable, Gratton, & Fabiani, 2003; Vuust et al., 2005; Winkler & Schröger, 1995). Can the MMN be used as individual marker of (auditory) temporal processing efficiency? Interindividual variability in brain structure and function has been shown to affect auditory processing mechanisms (see Zatorre, 2013 for a review). Previous studies reported correlations between MMN characteristics and discrimination accuracy on sound deviations on complex spectro-temporal patterns (e.g., Baldeweg, Richardson, Watkins, Foale, & Gruzelier, 1999; Lang et al., 1995; Näätänen et al., 1993; Sams, Paavilainen, Alho, & Näätänen, 1985; Tervaniemi et al., 2001). Similar findings were reported in the temporal domain: larger amplitude and earlier latencies reflect a better and faster detection of temporal deviations from standard sequences (Kujala et al., 2001; Ladinig, Honing, Hááden, & Winkler, 2009; Tervaniemi, Ilvonen, et al., 1997). Together, these suggest a neural processing disposition (not necessarily innate; Bigand & Poulin-Charronnat, 2006) by individuals for processing complex spectro-temporal patterns.

ERP latency seems to reflect learning or maturational states of the relevant neural networks: the degree of myelination of white-matter connectivity between cortical generators (Baggio, 2012; Cardenas et al., 2005) while ERP amplitude, seems to rely on the neuronal population size of cortical generators and the extent of active tissue (Näätänen & Alho, 1997). Myelination and functional connectivity between the relevant cortical generators may contribute to how fast auditory temporal mechanisms (Koelsch & Siebel, 2005) are for processing rapidly incoming information, and consequently, for detecting standard-deviant differences. The evidence presented in this section suggests that inter-subjects variability in MMN latency, as found in this study, likely reflects differences in (pre-attentive) information processing rather than global or contextual, non-functional factors. Early sensory memory mechanisms may then account for individual behavior performance observed in signaling games (Tiitinen et al., 1994).

4.4.3. Sequence complexity and behavioral prediction

Temporal regularity optimizes temporal integration mechanisms (Schwartze, Rothermich, Schmidt-Kassow, & Kotz, 2011; Tavano, Widmann, Bendixen, Trujillo-Barreto, & Schröger, 2014). Small temporal deviations are more easily detected in isochronous than irregular temporal patterns (Drake & Botte, 1993). Irregular temporal patterns may increasingly tax the auditory system, reducing the quality of memory templates. In turn, this would be echoed by slower and weaker responses from the pre-attentive change-detection system. The present results support this view. We found the weakest MMN response for the high-entropic stimuli. Noticeably, only the MMN latency generated by these stimuli was consistently predictive for behavioral performance. In line with Lumaca and Baggio (2016), we propose that more demanding auditory processes may be necessary for the emergence of inter-subject variability in behavior. If as noted before, high-entropic stimuli require more sensitive auditory processing mechanisms, even small differences in processing capabilities across individuals, as reflected by MMNs, would manifest in behavior (Kujala et al., 2001). This difference would be minimal for control and low-entropic stimuli, where temporal information is highly predictable (van Zuijen et al., 2004). Complexity reveal to be a critical ingredient to investigate individual information processing capacities.

4.4.4. Neural predictors of cultural behavior

Faithful transmission is one property of cumulative cultural evolution (Caldwell & Millen, 2008). This may depends on the difficulty of the cultural repertoire to be transmitted and the learning skills of the receiver (Sawa, 2002). An inaccurate learner will be an unreliable transmitter in the future. Here, we show that the 'timing' of the brain processes can distinguish good from bad learners, and consequently, accurate from inaccurate transmitters. In line with Lumaca & Baggio, (2016) we found longer MMN latencies for inaccurate learners: the ones showing low coordination, and inaccurate transmission within (low consistency) and between games (low transmission and high innovation). Changes in MMN amplitude and latency have been shown to predict variations in reaction times and hit-rates on auditory discrimination tasks (Aaltonen, Eerola, Lang, Uusipaikka, & Tuomainen, 1994; Atienza, Cantero, & Dominguez-Marin, 2002; Kujala et al., 2001; Tiitinen et al., 1994). Here, we show that early sensory mechanisms can also drive more complex, reproduction tasks. This result further supports and strengthen the memory interpretation of the MMN (Näätänen et al., 2007).

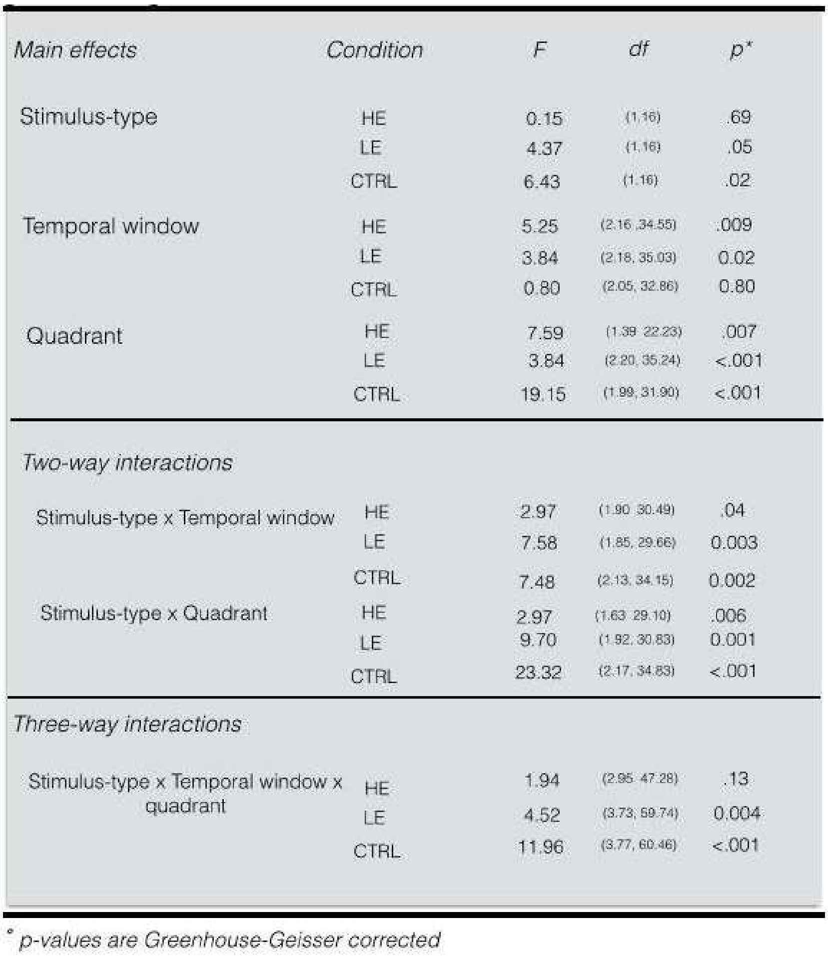
Perhaps more interesting is that 'timing' of early brain processes predicts the temporal regularization of the signaling system. Major properties of human symbolic systems, and cultural transmission systems in general, can be understood as adaptations to human brain "information bottleneck" (Kirby, 2001). The size of the information bottleneck determines properties of the symbolic system transmitted therein. Previous work shows that the tighter the bottleneck (i.e., the smaller the size of the data available to the learner), the more regular is the system in time (Kirby, Dowman, & Griffiths, 2007). In simulation studies, agents have no memory limits or learning constraints of any type. Here we suggest that similar results are obtained with tighter 'memory' bottleneck: the less efficient are the information processing capabilities of individuals, the more sensitive the same individuals are to pressures for learnability and regularization. Experiments on cultural research support this hypothesis (Kempe, Gauvrit, & Forsyth, 2015). In this experiment, the authors compared transmission chains of children with adults, the participant's role being to reproduce by memory the original positioning of visual-spatial patterns of dots. The random patterns of the first generations gradually increased in regularity in both groups as result of iterated transmission, but more readily in children. Cognitive demands to reproduce complex patterns can be higher for children, which have not yet fully developed their cognitive potential. This may explain why children injected more structure than adults into the code. In a fieldwork study, Dale and Lupian (2012) showed that American populations living in areas with high prevalence of nonnative speakers, tend to adopt more regular than irregular verbs. For nonnative adult speakers it can be hard to learn (idiosyncratic) irregular forms from another language. This may result in greater pressures to remove unpredictable variation, and to adopt regular forms. In this study, we support the view that complexity of the learning material taxes differently the human brain, according to its neural information processing capabilities, determining different pressures for the emergence of structure in cultural system. The greater the individual noise in perception and cognition (Spike, Smith, & Kirby, 2016), the larger the magnitude of adaptation by cultural material.

4.4.5. Individual processing constraints, cultural evolution and diversity

Cultural transmission is a powerful regularizer of symbolic systems (Deacon, 1997). Regularities emerge as a consequence of adaptation to structure and function of human cognition (Boyd & Richerson, 1988; Christiansen & Chater, 2008). Empirical works on language evolution support this view (e.g., Kirby, 2001; Kirby et al., 2008, 2007; Verhoef, 2012). However these studies have so far only focused on general pattern of changes (Caldwell et al., 2016), and the sources of these changes only inferred at population-level. Individual variability is there treated as *noise*, and removed by simple arithmetical procedures. Here we show that this variability, instead of representing random noise, may reflect the information processing capabilities of individuals. How does this variability manifest in the population structure? Through what mechanisms? Dediu and Ladd (2007) nominated cultural transmission as a potential one. Their proposal is that individual neural variability manifests in phenotypic differences in the acquisition and production of linguistic tones. This variability is amplified by cultural transmission to produce variation in linguistic patterns observed among world cultures in the use of linguistic tones. In this work, we further support this view in the laboratory (e.g., Kirby et al., 2008; Kirby et al., 2015; Ravignani et al., 2016), showing that small (and possibly different) individual biases on the acquisition, transmission, and regularization of rhythmic material, are amplified during cultural transmission in a diffusion chain, and manifest in the form of a design feature of music: isochronicity. Different evolutionary mechanisms might be also responsible for this outcome, such as content-based biases (Mesoudi, 2015). After the emergence of isochronicity, which could take longer in populations of good learners, the stimuli are easier to learn and recall. Individuals might have preferences to select these variants, which in turn, would spread widely in the population. Mechanisms of cultural evolution may provide a bridge to link individual neural variability and population-level patterns observed in cultural symbolic systems. The use of microsocieties, with the support of computational methods, can shed further lights on this matter. One way is to replicate this experiment with larger population of individuals. We could examine how isochronicity emerges at first, through the method of neural predictors, and how it spreads in communities (Tamariz, 2017). We could investigate to what extent populations of good-learners are affected by the ‘invasion’ (i.e., the introduction in the microsociety) of bad-learners, and much more.

The present work provides further support for a link between neural and cultural diversity (Lumaca & Baggio, 2016), and introduces a first experimental connection between neural variability and population structure (Dediu & Ladd, 2007). Phylogenetic methods constituted up to now the main approaches to explain cultural variation (Mesoudi, 2007; Savage & Brown, 2007). Changes over time in the frequency of cultural variants may produce peculiar population-pattern phenomena (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981), with a rich variability within the same culture (Rzeszutek et al., 2012). However, the source of the initial variability is often vague and assumed *a priori*. We propose that cognitive variation is a non-negligible source of cultural variation, supplementary to socio-environmental factors (Pagel & Mace, 2004). In the present context, it may have contributed to some forms of variation observed for temporal regularity over time and across cultures (Levitin et al., 2012; Nettl, 2000; Dediu & Ladd, 2007). Critically, this work shows a direct, visible trace of the relation between neural processes and cultural adaptation, up to now only assumed by cultural transmission research.

4.5. Figures and tables



Main effects	*Condition*	*F*	*df*	*p**
Stimulus-type	HE	0.15	(1,16)	.69
	LE	4.37	(1,16)	.05
	CTRL	6.43	(1,16)	.02
Temporal window	HE	5.25	(2.16 ,34.55)	.009
	LE	3.84	(2.18, 35.03)	0.02
	CTRL	0.80	(2.05, 32.86)	0.80
Quadrant	HE	7.59	(1.39 22.23)	.007
	LE	3.84	(2.20, 35.24)	<.001
	CTRL	19.15	(1.99, 31.90)	<.001
Two-way interactions				
Stimulus-type x Temporal window	HE	2.97	(1.90 30.49)	.04
	LE	7.58	(1.85, 29.66)	0.003
	CTRL	7.48	(2.13, 34.15)	0.002
Stimulus-type x Quadrant	HE	2.97	(1.63 29.10)	.006
	LE	9.70	(1.92, 30.83)	0.001
	CTRL	23.32	(2.17, 34.83)	<.001
Three-way interactions				
Stimulus-type x Temporal window x quadrant	HE	1.94	(2.95 47.28)	.13
	LE	4.52	(3.73, 59.74)	0.004
	CTRL	11.96	(3.77, 60.46)	<.001

° p-values are Greenhouse-Geisser corrected

Table 1. Results of ANOVA statistics on mean MMN amplitude values in the high entropy (HE), low entropy (LE) and control conditions (CTRL).

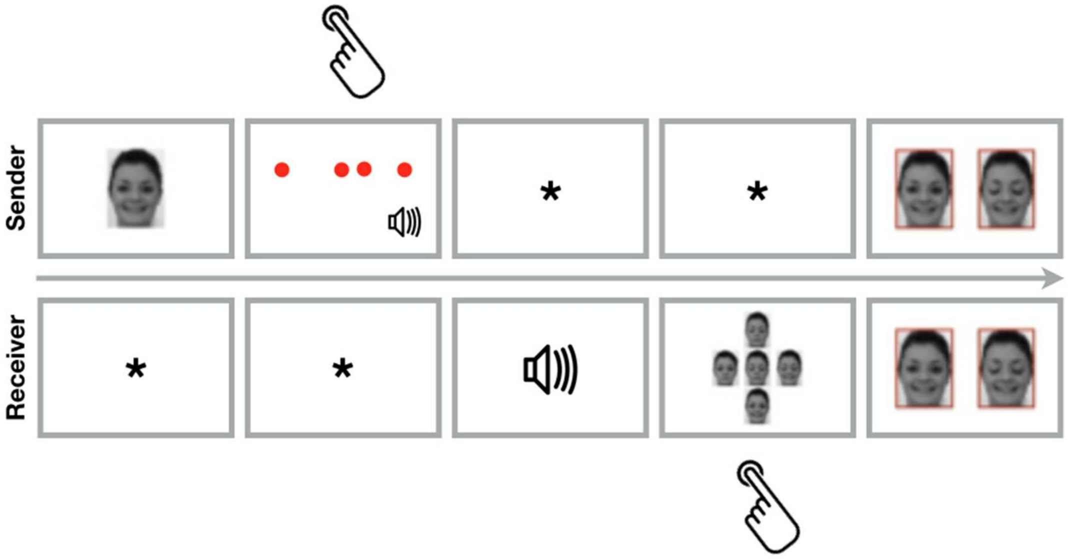


Fig. 1. Sample a trial from the signaling games played by participants in Experiments 1 and 2. The top and bottom rows show what senders and receivers saw on their screens, respectively. The task for the sender was to compose a 5-tone sequence to be used as a signal for the simple or compound emotion expressed by the face presented on the screen at the start of the trial, and for the receiver to respond to that signal by choosing the face the sender may have seen. The sender and the receiver converged over trials on a shared mapping of signals (monotone sequences) to meanings (emotions). Hand symbols indicate when the sender and the receiver had to produce a response. Feedback was provided to both players simultaneously, displaying the face seen by the sender and the face selected by the receiver in a green frame (matching faces; correct) or in a red frame (mismatching faces; incorrect). Time flows from left to right.

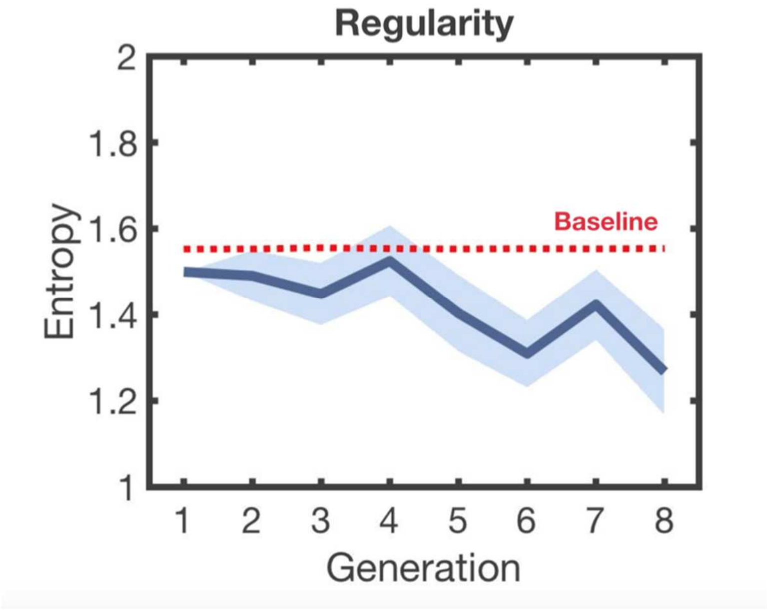


Fig. 2. Evolution of rhythmic regularity across generations in Experiment 1. The shaded blue area shows standard errors of the mean (SEM).

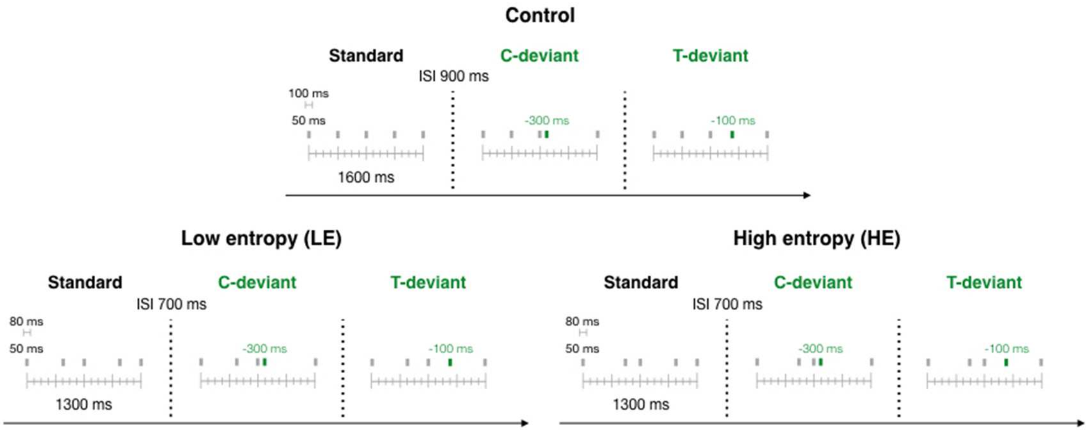


Fig. 3. Schematic illustration of the rhythm sequences presented to participants during day 1 of Experiment 2. The EEG was recorded while participants listened to these rhythm sequences. Each sequence consisted of 5 tones occurring at varying temporal intervals from one another in different conditions (standard, C-deviant or T-deviant). Sequences of different complexity were used: control (isochrone) sequences, low (LE) and high (HE) entropy sequences. The proportion of standards to deviants was 4:1. A blocked design was used, where the blocks were randomized across participants.

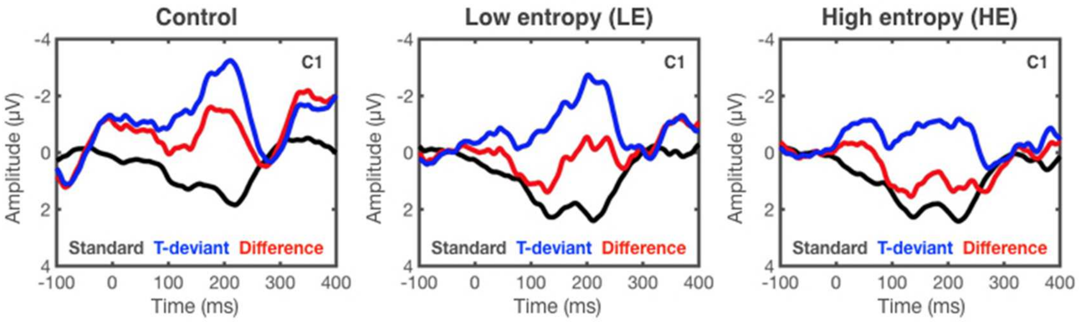


Fig. 4. Grand-average ERPs (N=17) from a left central site (C1) in response to the 4th tone in standard and T-deviant stimuli in control, low and high entropy sequences. The red lines show difference waves between T-deviants and standards. Waveforms were low-pass filtered at 30Hz.

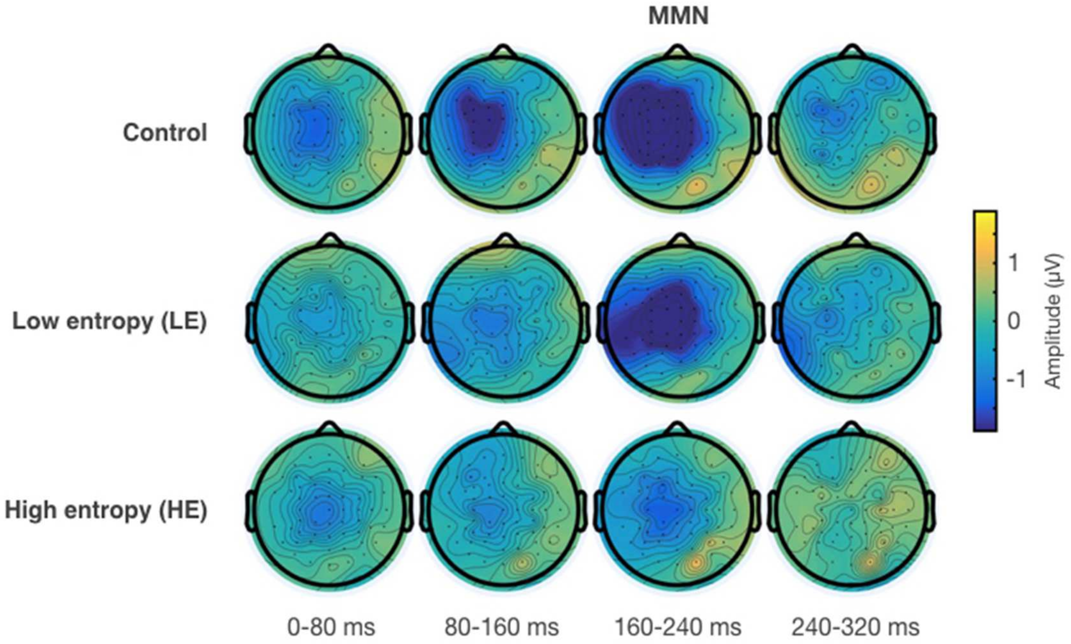


Fig. 5. Topographic isovoltage maps of grand-average (N=17) ERP difference waves between T-deviants and standards in control, low and high entropy sequences. The mismatch negativity (MMN) corresponds to the topographic maps shown at 160-240 ms; 0 ms is the onset of the 4th tone in a sequence.

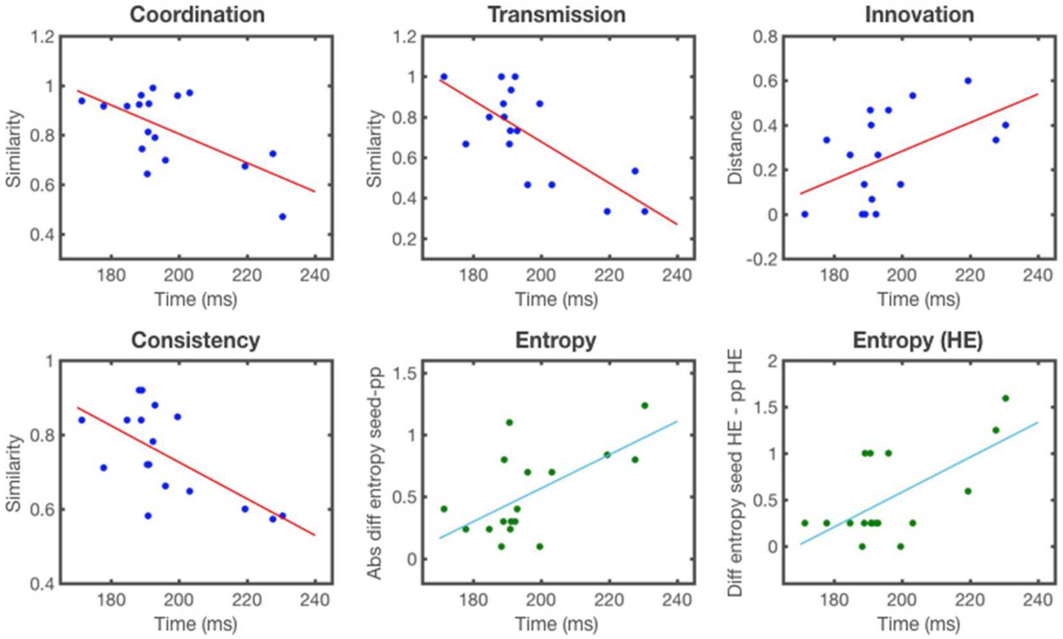


Fig. 6. Pearson product-moment correlations (*r*) between MMN peak latencies relative to the onset of the deviant tone (0 ms) in high entropy sequences from the left hemispheric scalp quadrant on day 1 (*x*-axis), and behavioral (blue, red) and structural (green, light blue) measures in signaling games on day 2 (*y*-axis). Structural measure was obtained as the absolute difference between mean values in Games 1 and 2. All MMN peaks fall within the 160-240 ms window. See Methods for further details on the measures, and Suppl. Tabs. 2.3.1-2.3.2 for Pearson product-moment correlations (*r*) values. Each point on a scatterplot is one participant (n=17).

4.6. Supplementary material

1. Experiment 1

1.1. Written instructions to participants

Before the signaling games (SGs) session, each participant was informed he would take part in two successive games, each played in a different role. In Game 1, he would be the receiver. He was asked to try to learn which facial expressions (emotions) corresponded to the 5-equitone sequences sent by the other player (sender). There was no reward for speed, either in a trial or in the game. Successful coordination was the only reward for participants. In Game 2, the participant played as sender. He was instructed to transmit the tone system acquired in Game 1, as he remembered it, to a new receiver.

1.2. Additional structural measures

1.2.1. Metric complexity

Our measure of *metric complexity* calculates the number of hierarchical levels with chunked pairs of subdivisions of the main beat (or tactus) required to match each onset in one rhythm. Additional calculations based on this estimate are applied to find the meter of a rhythm and to calculate the degree of syncopation. First, we calculate the number of subdivisions of the tactus necessary to fit the rhythm for each note onset (at a precision level equal to the smallest quantization unit divided by two). Second, the metric complexity of each onset in a rhythm is estimated by rounding down to the nearest integer value the log2 (base 2 logarithm) of the required number of subdivisions. Last, the metric complexity of the rhythm is estimated as the sum of the metrical complexities of each note onset in the rhythm. Then, a principle of economy has been used to find the correct meter. We assume that the perceptually most valid meter is the simplest one. It follows that the meter for a rhythm can be estimated from a set of predefined possible meters by finding the regularly spaced main beats or tactus of the meter (e.g., binary meter with four beats, ternary meter with three beats, or a quinary meter with five beats) which minimize the metric complexity value.

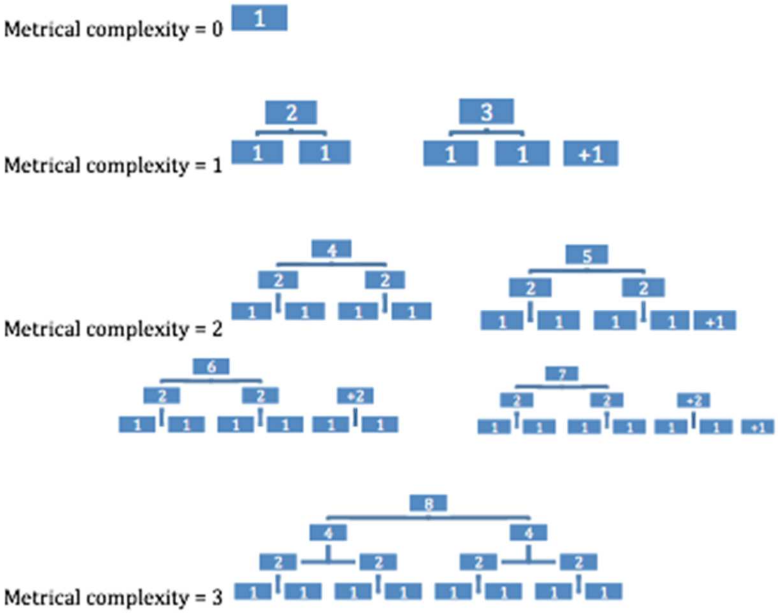


Figure 1.2.1. Diagram-illustration of metrical hierarchies for subdivisions of the main beat or tactus chunked into binary (i.e., base 2) relationships plus eventual remainder single elements. The total number of subdivisions is shown at the top of the metrical hierarchy. The constituents' binary elements are shown for each chunk level. The number of levels determines the metric complexity value.

1.2.2. Syncopation

A measure of syncopation similar to Witek et al. (2014) was used. This measure is based on the metric complexity estimate. For each note onset, if the following note onset occurs after the next main beat or tactus, we assign a syncopation value that correspond to the difference in metric complexity between the note onset and the main beat or tactus (metric complexity value - 0). Otherwise a value of 0 is assigned. The syncopation value of the rhythm is the sum of the syncopation values of each note onset.

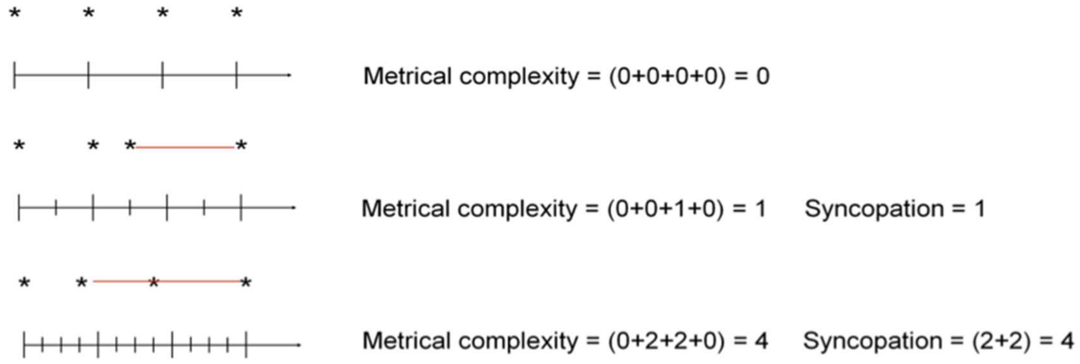


Figure 1.2.2. Examples on how to calculate the syncopation of different rhythms based on metric complexity.

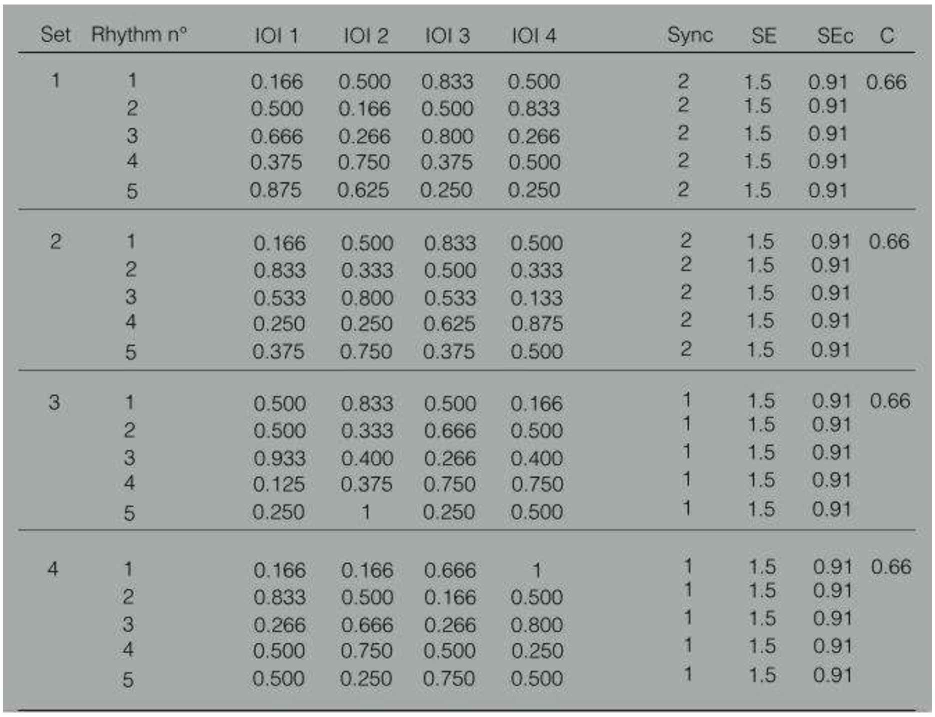
1.2.3. Compression

Rhythmic compression measures how similar are rhythms of the same signal set (Cornish et al., 2013). It was computed by taking the mean similarity between the contour of rhythms of the same signal set. It range from 0 (different rhythmic contour profiles) to 1 (mono-contour signal set).

1.2.4. Compositionality

Compositionality measures the degree of systematic relation between signals and meanings. Compositionality of the rhythmic set was computed with the information-theoretic tool *RegMap* (for details see Cornish et al., 2009). For the purpose of this measure, upper face of states was coded using two variants (*open* and *closed* eyes); the lower face involved three variants (mouth corners *up, straight* and *down*). Upper and lower face corresponds to the meaning dimensions. Then, the entire pool of signals was partitioned into segments, in all possible combinations: units of 3 and 1 IOIs (3+1), in units of 2 and 2 IOIs (2+2), and finally in units of 1 and 3 IOIs (1+3). For each of these partitions, the conditional entropy of any possible combination of signal-meaning dimensions was computed under 1000 randomizations, obtaining a z-partial RegMap. This measure specifies to what degree a signal unit reliably codes a meaning dimension. Also, a single z-RegMap value for the entire code was computed (range 0:1) which summarize the compositionality of the set of signals produced by the pair. Results are shown in Suppl. Fig. 1.4.5.

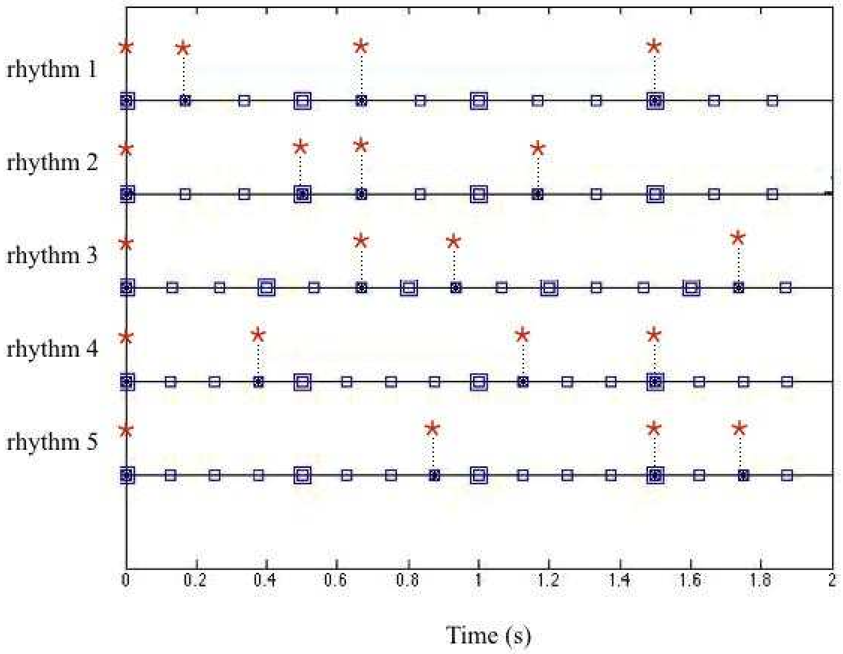
1.3. Supplementary Tables



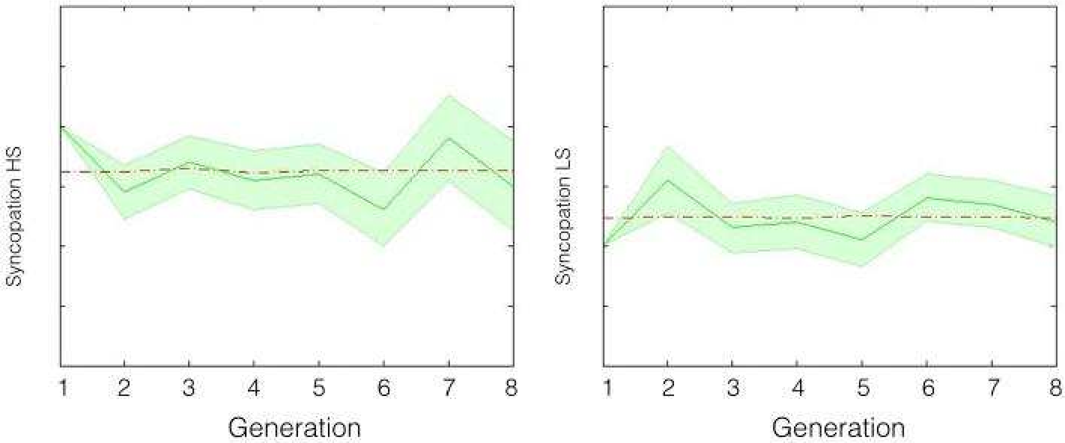
Set	Rhythm n°	IOI 1	IOI 2	IOI 3	IOI 4	Sync	SE	SEc	C
1	1	0.166	0.500	0.833	0.500	2	1.5	0.91	0.66
	2	0.500	0.166	0.500	0.833	2	1.5	0.91	
	3	0.666	0.266	0.800	0.266	2	1.5	0.91	
	4	0.375	0.750	0.375	0.500	2	1.5	0.91	
	5	0.875	0.625	0.250	0.250	2	1.5	0.91	
2	1	0.166	0.500	0.833	0.500	2	1.5	0.91	0.66
	2	0.833	0.333	0.500	0.333	2	1.5	0.91	
	3	0.533	0.800	0.533	0.133	2	1.5	0.91	
	4	0.250	0.250	0.625	0.875	2	1.5	0.91	
	5	0.375	0.750	0.375	0.500	2	1.5	0.91	
3	1	0.500	0.833	0.500	0.166	1	1.5	0.91	0.66
	2	0.500	0.333	0.666	0.500	1	1.5	0.91	
	3	0.933	0.400	0.266	0.400	1	1.5	0.91	
	4	0.125	0.375	0.750	0.750	1	1.5	0.91	
	5	0.250	1	0.250	0.500	1	1.5	0.91	
4	1	0.166	0.166	0.666	1	1	1.5	0.91	0.66
	2	0.833	0.500	0.166	0.500	1	1.5	0.91	
	3	0.266	0.666	0.266	0.800	1	1.5	0.91	
	4	0.500	0.750	0.500	0.250	1	1.5	0.91	
	5	0.500	0.250	0.750	0.500	1	1.5	0.91	

Supplementary Table 1.3. Seeding material and relative structural properties. Four signal sets (1-4) of 5 temporal patterns, were used as the starting material for the diffusion chains. We used high-syncopated (sets 1-2) and low-syncopated stimuli (sets 3-4). Time of inter-onset intervals (IOIs) is shown in seconds. *Abbreviations:* IOI (inter-onset interval); Sync (syncopation); SE (Shannon Entropy); SEc (Shannon Entropy on rhythmic contour); C (mean compression on rhythmic contour). Sound files (enclosed) provide examples of the temporal patterns.

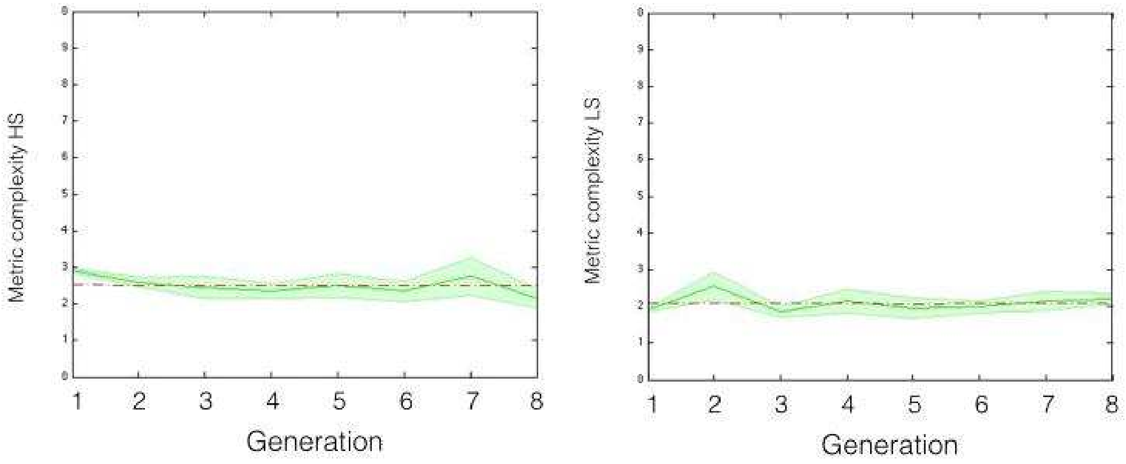
1.4. Supplementary Figures



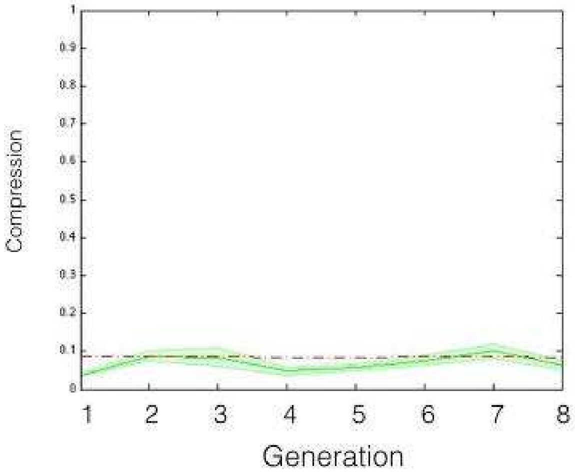
Supplementary Figure 1.4.1. Diagram-notation of the temporal patterns in the seed Set 1. Rhythms are ordered from the top (rhythm 1) to bottom. Red asterisks show the note onsets. Blue squares represent the metrical grid of reference. Time (in seconds) flows from left to right.



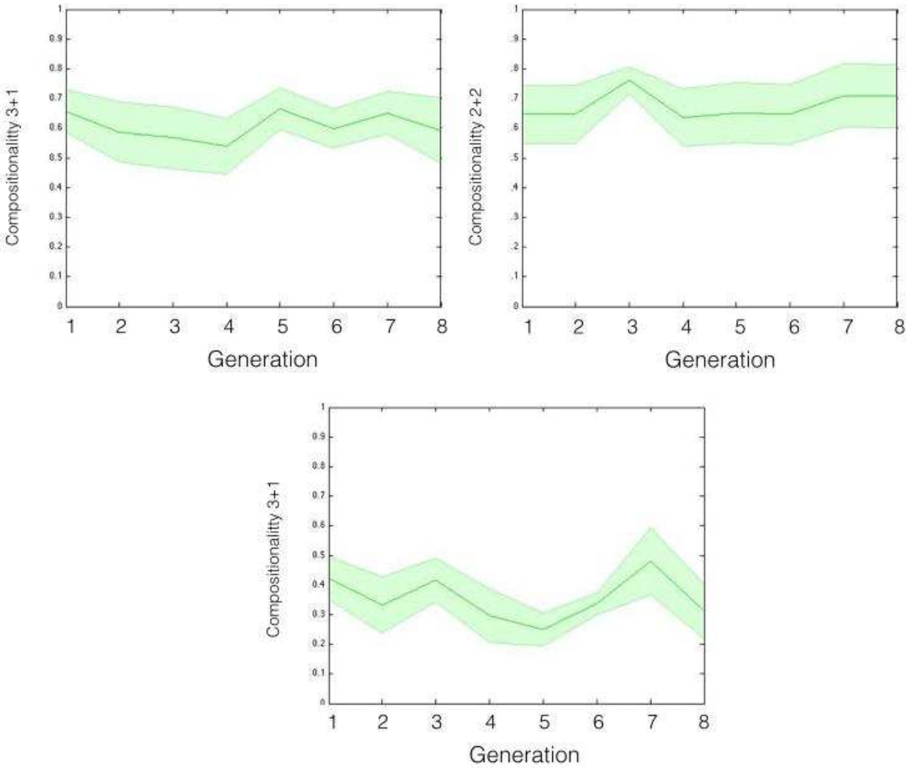
Supplementary Figure 1.4.2. Evolution of ***syncopation*** across generations for high-syncopated (HS; left) and low-syncopated set stimuli (LS; right). No significant changes over time were observed between the first and the last generation. No differences with the baseline were reported in the last generation.



Supplementary Figure 1.4.3. Evolution of ***metric complexity*** across generations for high-syncopated (HS; left) and low-syncopated set stimuli (LS; right). No significant changes over time were observed between the first and the last generation. No differences with the baseline were reported in the last generation.



Supplementary Figure 1.4.4. Evolution of **mean *compression*** across generations. No significant changes were observed between the first and the last generation. No differences with the baseline were reported in the last generation.



Supplementary Figure 1.4.5. Evolution of ***compositionality*** across generations when signals were segmented in 3+1 tone segments (*upper left*), 2+2 tone segments (*upper right*) and 1+3 tone segments (*bottom*). For details see Suppl. Inf. 1.1.2.4. No significant changes were observed between the first and the last generation.

2. Experiment 2

2.1. Written instructions to participants

2.1.1. ***Day* 1.** During the EEG session, the participant was asked to watch a silent movie without paying any attention to the temporal patterns played through the loudspeakers; the movie was subtitled in English. To direct their attention away from the tone sequences, we told participants that there would be questions about the movie at the end of the session.

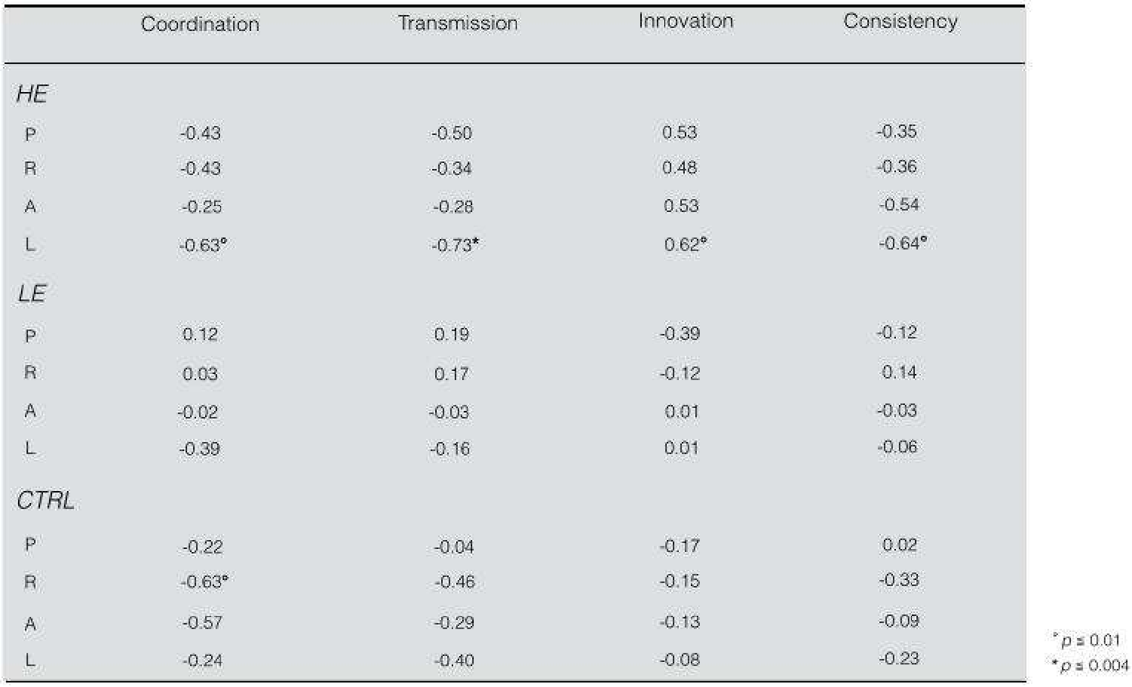
2.1.2. ***Day* 2.** See Suppl. Inf. 1.1.

2.2. Confederate behavior

2.2.1. ***Game* 1.** The confederate was trained on a particular mapping of emotions to keys on the keyboard (see Procedure, main article): in Game 1, he played with no sound feedback (volume set at 0), to prevent him from learning the ‘seed’ system and be biased in Game 2. The confederate was instructed to use consistently that mapping throughout Game 1, regardless of what the participant (receiver) would do.

2.2.2. ***Game* 2**. The confederate played according to the same written instructions given to the participant (receiver) in Game 1, namely, he was instructed to try to learn the temporal patterns as produced by the participant (sender).

2.3. Supplementary Tables



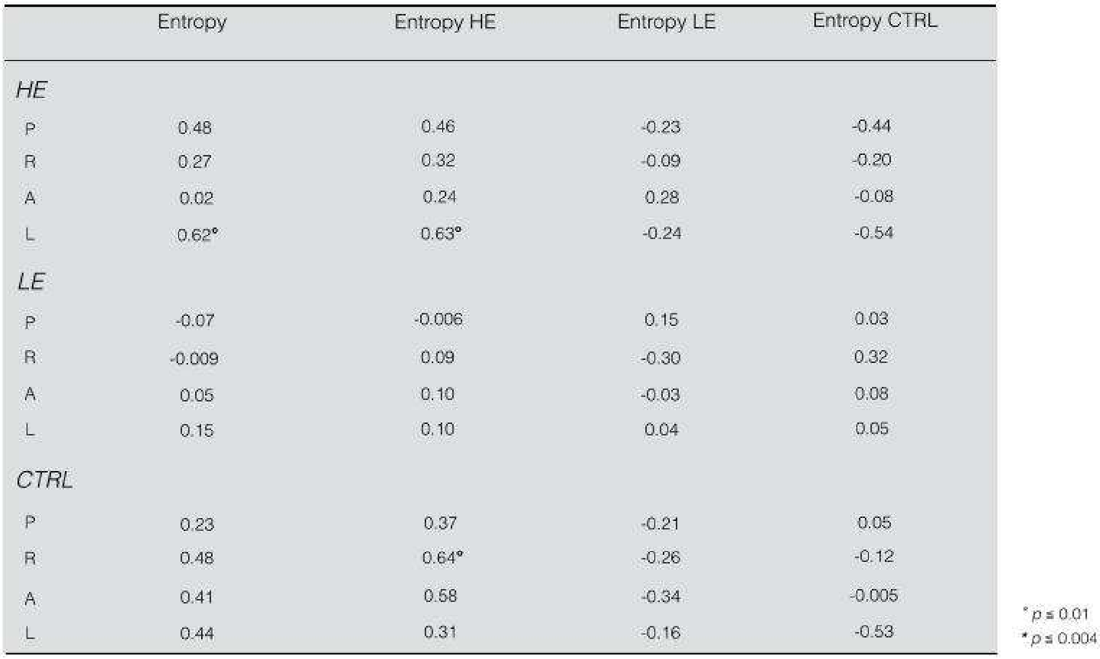
	Coordination	Transmission	Innovation	Consistency
HE				
P	-0.43	-0.50	0.53	-0.35
R	-0.43	-0.34	0.48	-0.36
A	-0.25	-0.28	0.53	-0.54
L	-0.63°	-0.73*	0.62°	-0.64°
LE				
P	0.12	0.19	-0.39	-0.12
R	0.03	0.17	-0.12	0.14
A	-0.02	-0.03	0.01	-0.03
L	-0.39	-0.16	0.01	-0.06
CTRL				
P	-0.22	-0.04	-0.17	0.02
R	-0.63°	-0.46	-0.15	-0.33
A	-0.57	-0.29	-0.13	-0.09
L	-0.24	-0.40	-0.08	-0.23

° $p \leq 0.01$
* $p \leq 0.004$

Supplementary Table 2.3.1. Pearson product-moment correlations (*r*) between MMN ***peak latencies*** at 160-240[4] ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***behavioral measures*** (coordination, transmission, innovation, consistency). Coefficients are shown for each scalp quadrant (where P is posterior, R is central-right, A is anterior, and L is central-left). P-values are Bonferroni corrected at $\alpha=0.05/12=0.004$, where 12 is the number of tests involving each independent variable.

---

[4] In all cases, the MMN peak fell within the 160-240 ms window. See main article, Fig. 4.



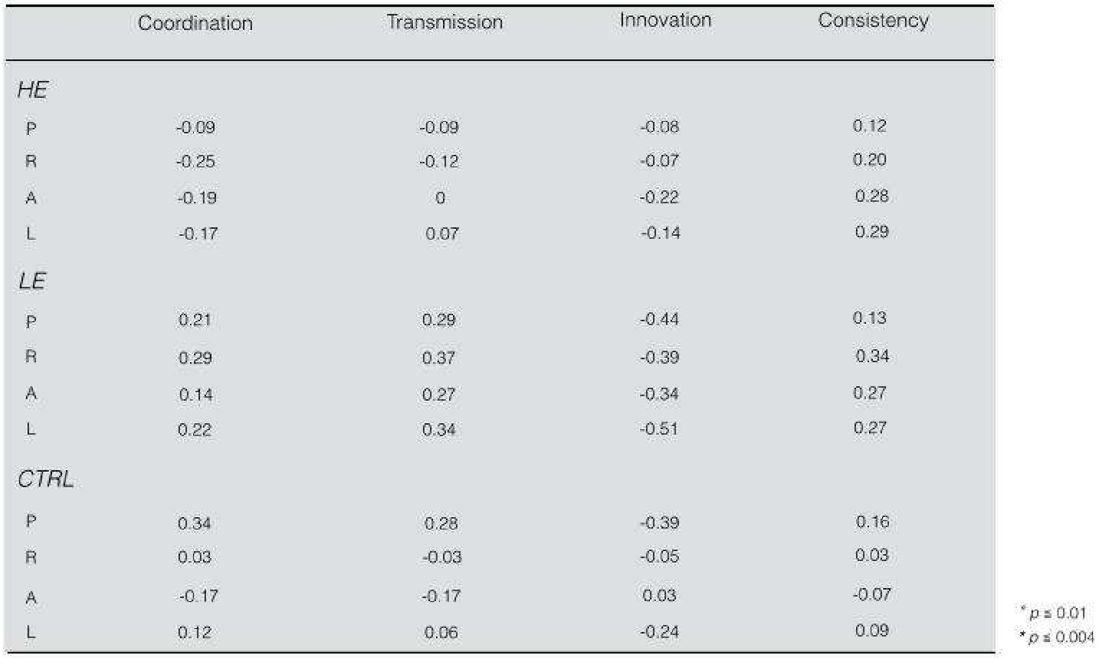
	Entropy	Entropy HE	Entropy LE	Entropy CTRL
HE				
P	0.48	0.46	-0.23	-0.44
R	0.27	0.32	-0.09	-0.20
A	0.02	0.24	0.28	-0.08
L	0.62°	0.63°	-0.24	-0.54
LE				
P	-0.07	-0.006	0.15	0.03
R	-0.009	0.09	-0.30	0.32
A	0.05	0.10	-0.03	0.08
L	0.15	0.10	0.04	0.05
CTRL				
P	0.23	0.37	-0.21	0.05
R	0.48	0.64°	-0.26	-0.12
A	0.41	0.58	-0.34	-0.005
L	0.44	0.31	-0.16	-0.53

° $p \leq 0.01$
* $p \leq 0.004$

Supplementary Table 2.3.2. Pearson product-moment correlations (*r*) between MMN ***peak latencies*** at 160-240[5] ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***structural measures*** (absolute difference entropy pp-seed on the *left*; difference HE, LE, CTRL codes - relative codes in the seed on the *right*). Coefficients are shown for each scalp quadrant (where P is posterior, R is central-right, A is anterior, and L is central-left). P-values are Bonferroni corrected at □=0.05/12=0.004, where 12 is the number of tests involving each independent variable.

---

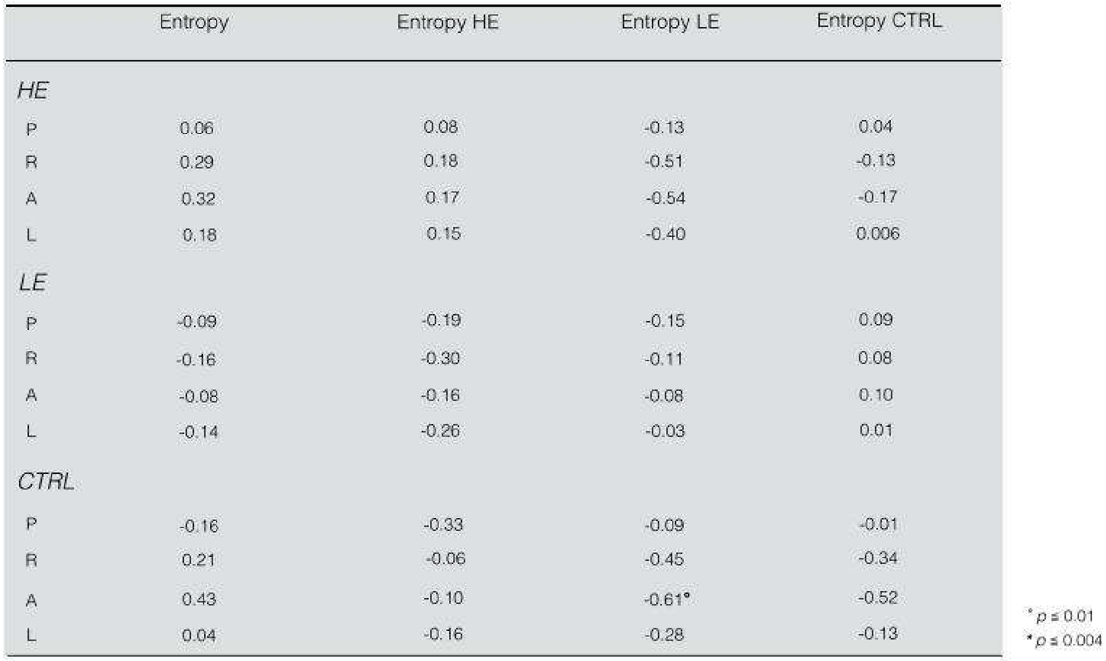
[5] In all cases, the MMN peak fell within the 160-240 ms window. See main article, Fig. 4.



	Coordination	Transmission	Innovation	Consistency
HE				
P	-0.09	-0.09	-0.08	0.12
R	-0.25	-0.12	-0.07	0.20
A	-0.19	0	-0.22	0.28
L	-0.17	0.07	-0.14	0.29
LE				
P	0.21	0.29	-0.44	0.13
R	0.29	0.37	-0.39	0.34
A	0.14	0.27	-0.34	0.27
L	0.22	0.34	-0.51	0.27
CTRL				
P	0.34	0.28	-0.39	0.16
R	0.03	-0.03	-0.05	0.03
A	-0.17	-0.17	0.03	-0.07
L	0.12	0.06	-0.24	0.09

° $p \leq 0.01$
* $p \leq 0.004$

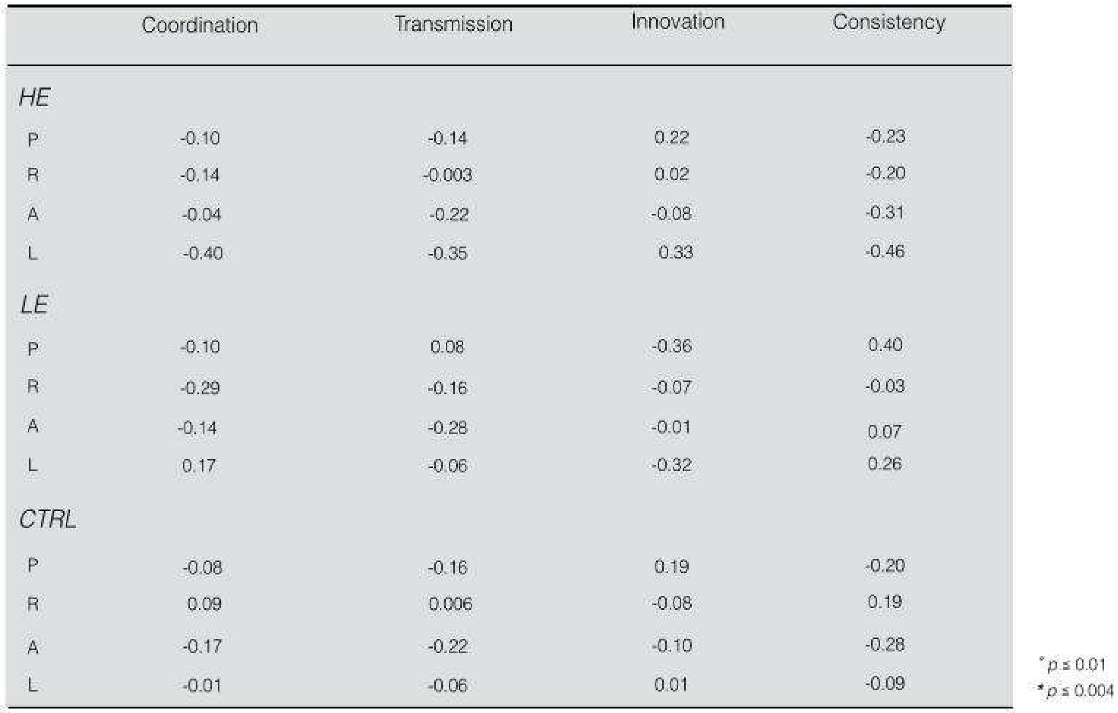
Supplementary Table 2.3.3. Pearson product-moment correlations (r) between MMN ***mean amplitudes*** at 160-240 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***behavioral measures*** (coordination, transmission, innovation, consistency). Coefficients are shown for each scalp quadrant (where P is posterior, R is central-right, A is anterior, and L is central-left). P-values are Bonferroni corrected at =0.05/12=0.004, where 12 is the number of tests involving each independent variable.



	Entropy	Entropy HE	Entropy LE	Entropy CTRL
HE				
P	0.06	0.08	-0.13	0.04
R	0.29	0.18	-0.51	-0.13
A	0.32	0.17	-0.54	-0.17
L	0.18	0.15	-0.40	0.006
LE				
P	-0.09	-0.19	-0.15	0.09
R	-0.16	-0.30	-0.11	0.08
A	-0.08	-0.16	-0.08	0.10
L	-0.14	-0.26	-0.03	0.01
CTRL				
P	-0.16	-0.33	-0.09	-0.01
R	0.21	-0.06	-0.45	-0.34
A	0.43	-0.10	-0.61°	-0.52
L	0.04	-0.16	-0.28	-0.13

° $p \le 0.01$
* $p \le 0.004$

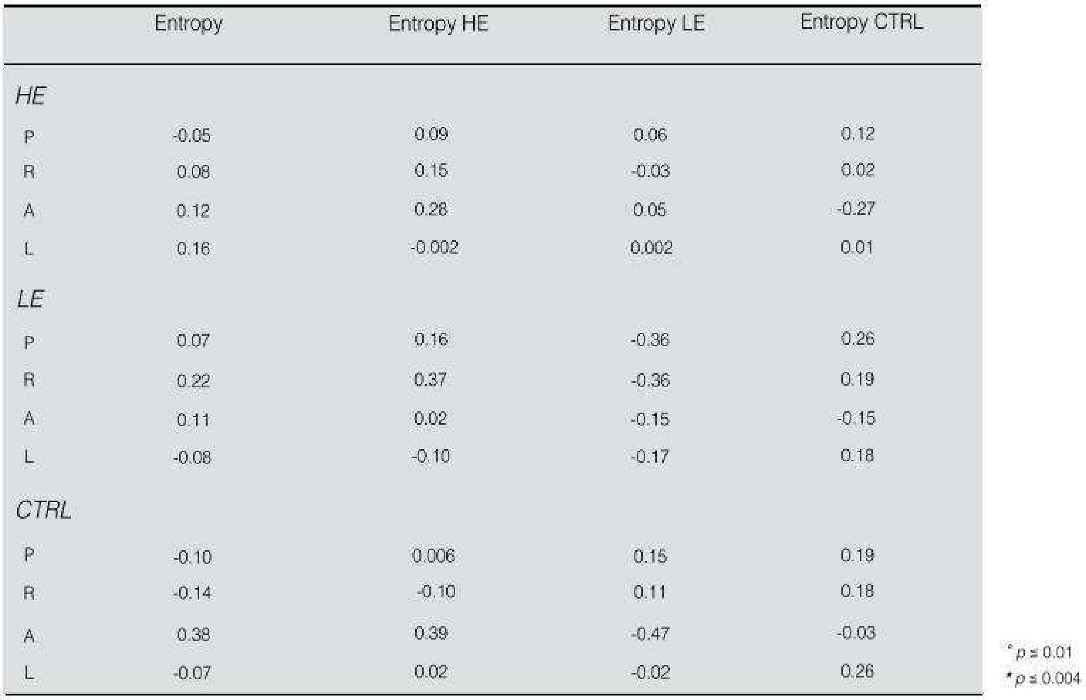
Supplementary Table 2.3.4. Pearson product-moment correlations (*r*) between MMN ***mean amplitudes*** at 160-240 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***structural measures*** (absolute difference entropy pp-seed on the *left*; difference HE, LE, CTRL codes - relative codes in the seed on the *right*). Coefficients are shown for each scalp quadrant (where P is posterior, R is central-right, A is anterior, and L is central-left). P-values are Bonferroni corrected at □=0.05/12=0.004, where 12 is the number of tests involving each independent variable.



	Coordination	Transmission	Innovation	Consistency
HE				
P	-0.10	-0.14	0.22	-0.23
R	-0.14	-0.003	0.02	-0.20
A	-0.04	-0.22	-0.08	-0.31
L	-0.40	-0.35	0.33	-0.46
LE				
P	-0.10	0.08	-0.36	0.40
R	-0.29	-0.16	-0.07	-0.03
A	-0.14	-0.28	-0.01	0.07
L	0.17	-0.06	-0.32	0.26
CTRL				
P	-0.08	-0.16	0.19	-0.20
R	0.09	0.006	-0.08	0.19
A	-0.17	-0.22	-0.10	-0.28
L	-0.01	-0.06	0.01	-0.09

°$p \leq 0.01$
*$p \leq 0.004$

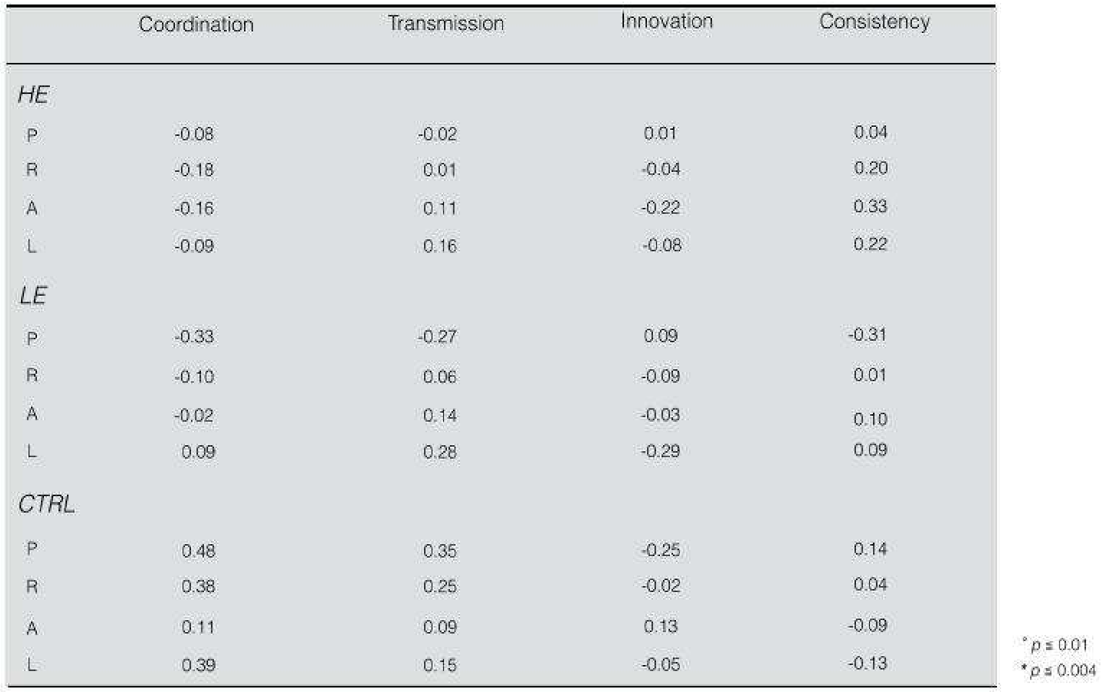
Supplementary Table 2.3.5. Pearson product-moment correlations (*r*) between MMN ***peak latencies*** at 240-320 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***behavioral measures*** (coordination, transmission, innovation, consistency). Coefficients are shown for each scalp quadrant (where P is posterior, R is central-right, A is anterior, and L is central-left). P-values are Bonferroni corrected at α=0.05/12=0.004, where 12 is the number of tests involving each independent variable.



	Entropy	Entropy HE	Entropy LE	Entropy CTRL
HE				
P	-0.05	0.09	0.06	0.12
R	0.08	0.15	-0.03	0.02
A	0.12	0.28	0.05	-0.27
L	0.16	-0.002	0.002	0.01
LE				
P	0.07	0.16	-0.36	0.26
R	0.22	0.37	-0.36	0.19
A	0.11	0.02	-0.15	-0.15
L	-0.08	-0.10	-0.17	0.18
CTRL				
P	-0.10	0.006	0.15	0.19
R	-0.14	-0.10	0.11	0.18
A	0.38	0.39	-0.47	-0.03
L	-0.07	0.02	-0.02	0.26

° $p \leq 0.01$
* $p \leq 0.004$

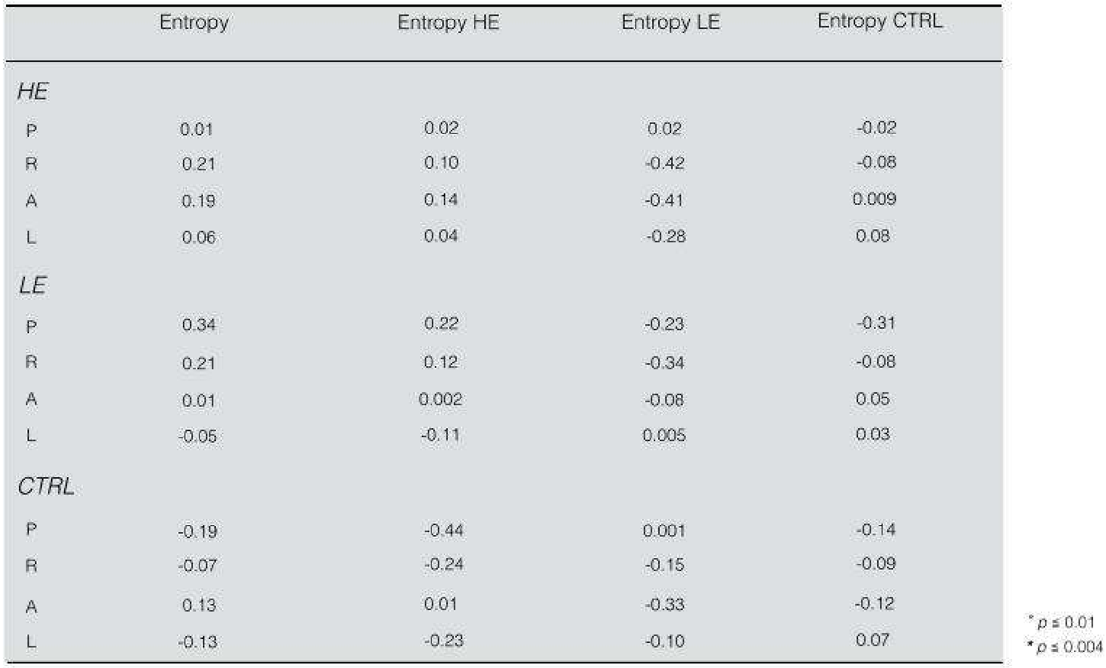
Supplementary Table 2.3.6. Pearson product-moment correlations (*r*) between MMN ***peak latencies*** at 240-320 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***structural measures*** (absolute difference entropy pp-seed on the *left*; difference HE, LE, CTRL codes - relative codes in the seed on the *right*). Coefficients are shown for each scalp quadrant (where P is posterior, R is central-right, A is anterior, and L is central-left). P-values are Bonferroni corrected at □=0.05/12=0.004, where 12 is the number of tests involving each independent variable.



	Coordination	Transmission	Innovation	Consistency
HE				
P	-0.08	-0.02	0.01	0.04
R	-0.18	0.01	-0.04	0.20
A	-0.16	0.11	-0.22	0.33
L	-0.09	0.16	-0.08	0.22
LE				
P	-0.33	-0.27	0.09	-0.31
R	-0.10	0.06	-0.09	0.01
A	-0.02	0.14	-0.03	0.10
L	0.09	0.28	-0.29	0.09
CTRL				
P	0.48	0.35	-0.25	0.14
R	0.38	0.25	-0.02	0.04
A	0.11	0.09	0.13	-0.09
L	0.39	0.15	-0.05	-0.13

° $p \leq 0.01$
* $p \leq 0.004$

Supplementary Table 2.3.7. Pearson product-moment correlations (r) between MMN ***mean amplitudes*** at 240-320 ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***behavioral measures*** (coordination, transmission, innovation, consistency). Coefficients are shown for each scalp quadrant (where P is posterior, R is central-right, A is anterior, and L is central-left). P-values are Bonferroni corrected at α=0.05/12=0.004, where 12 is the number of tests involving each independent variable.



	Entropy	Entropy HE	Entropy LE	Entropy CTRL
HE				
P	0.01	0.02	0.02	-0.02
R	0.21	0.10	-0.42	-0.08
A	0.19	0.14	-0.41	0.009
L	0.06	0.04	-0.28	0.08
LE				
P	0.34	0.22	-0.23	-0.31
R	0.21	0.12	-0.34	-0.08
A	0.01	0.002	-0.08	0.05
L	-0.05	-0.11	0.005	0.03
CTRL				
P	-0.19	-0.44	0.001	-0.14
R	-0.07	-0.24	-0.15	-0.09
A	0.13	0.01	-0.33	-0.12
L	-0.13	-0.23	-0.10	0.07

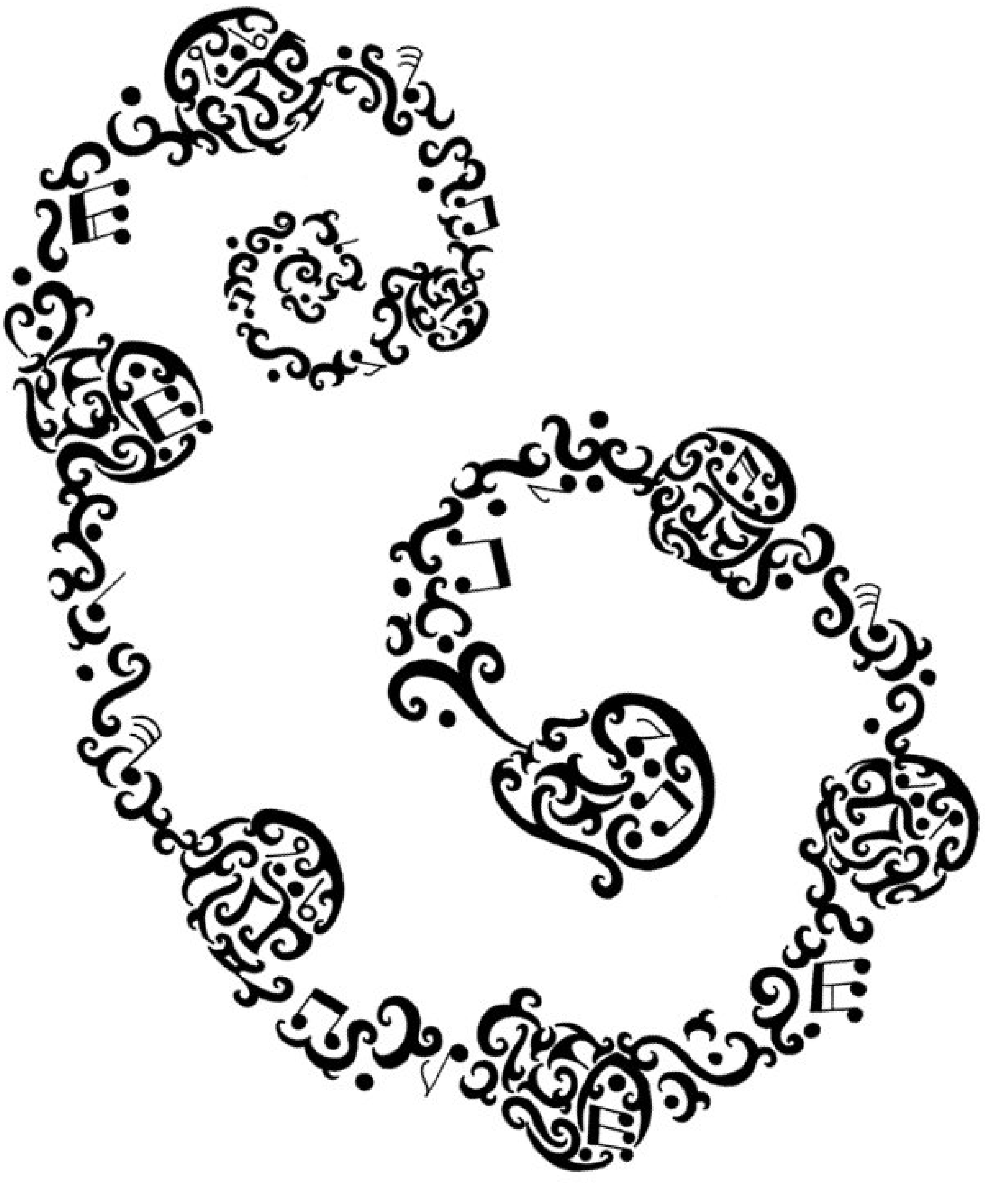
˚ $p \leq 0.01$

* $p \leq 0.004$

Supplementary Table 2.3.8. Pearson product-moment correlations (*r*) between MMN ***mean amplitudes*** at 240-320ms relative to the onset of the 4th-tone deviant (0 ms) in high entropy (HE), low entropy (LE) and control stimuli (CTRL), and ***structural measures*** (absolute difference entropy pp-seed on the *left*; difference HE, LE, CTRL codes - relative codes in the seed on the *right*). Coefficients are shown for each scalp quadrant (where P is posterior, R is central-right, A is anterior, and L is central-left). P-values are Bonferroni corrected at □=0.05/12=0.004, where 12 is the number of tests involving each independent variable.

Author Contributions

ML, NTH, GB conceived and designed the experiment. ML conducted research and analyzed the data. ML and GB wrote the paper. ML, NTH, GB wrote the Supplementary material. PV and EB co-supervised the experiment and critically commented on the design, data analyses and manuscript.



Chapter 5

Overview and general discussion

Massimo Lumaca

Abstract

In this thesis I have addressed in the laboratory the role of cognitive constraints on the cultural evolution of certain basic aspects of musical structure. I did this with a combination of behavioral and neurophysiological methods. First, I examined the evolution of structural regularities in (auditory) artificial symbolic systems while they were being transmitted across individuals, through interaction and coordination. Then, I explored the link between changes introduced by individuals and neural information processing. More specifically, I examined whether information processing capabilities of single individuals, as measured with event-related potentials (ERPs), could predict how they learned, transmitted and regularized an artificial symbolic system. I did this separately for two basic dimensions of musical structure, namely melody and rhythm. These studies support the idea that cognitive constraints on cultural transmission can account for some forms of variant and invariant aspects found within and across symbolic systems, music in particular. Perhaps more relevant, they provide initial neurophysiological evidence of adaptation by symbolic systems to constraints on information processing of the human brain.

5.1. Overview of the findings in this thesis

In the first empirical chapter (**chapter 2**) I investigated the evolution of melodic systems experimentally using a new laboratory model of cultural evolution: the multi-generational signaling games (or MGSGs). More specifically, I studied how artificial systems of tone sequences changed in the course of iterated signaling games, where each iteration introduces a new generation of performers and listeners. I observed the gradual evolution of several structural regularities of musical phrases that may promote learning and memory retrieval: pitch proximity (tendency for small pitch intervals), pitch continuity (tendency for a single melodic direction), melodic symmetry (melodic forms that mirror others) and melodic compression (tone sequences of the same signal set become increasingly similar). Crucially, most of these features are universal traits of musical systems. I argued that our experiment tapped into universal cognitive mechanisms of structured sequence processing, most probably at work also on the cultural evolution of auditory symbolic systems, music in particular. This study is the first to describe the emergence of structural properties of melodic systems in the laboratory, using an explicit lab-based model of cultural transmission in which agents interact and exchange information.

An interesting and open question emerges from this study. A common belief in cultural evolution research is that every cultural phenomenon can be understood as the population-level result of individual-level phenomena, in turn reflecting functional aspects of human cognition (Dediu & Ladd, 2007). Psychological properties of individuals would manifest, through different cultural dynamics, as structural properties of culture. Iterated learning research provided the first empirical support for the existence of this link (see **chapter 1**). A natural question is whether it is possible to move one level further: namely, would ‘neural’ properties of individuals manifest as structural properties of culture? And, as a consequence, may differences in neural properties across individuals determine variant aspects (across individuals and over time) of cultural symbolic system? Up to now, this level of explanation has remained unexplored.

Chapter 3 addressed this question, with the rationale that if population-level regularities originate in neural constraints, then it should be possible to predict individual regularization of cultural material on the base of neural data. As a follow up of the previous study, I focused on the melodic dimension. To test this hypothesis, I combined EEG and behavioral experiments in two successive days. I showed that the latency of the mismatch negativity (MMN) recorded in a pre-attentive auditory sequence processing task on day 1, could predict how well participants learned and transmitted an artificial symbolic system (similar to the one used in **chapter 2**) in two signaling games on day 2. Notably, MMN latencies also predicted which structural changes were introduced by participants into the artificial tone system. Specifically, individuals with longer MMN peak latencies were those that innovated the code more (i.e., they were less faithful transmitters), introducing more structural regularities, such as proximal intervals and smooth contours on the melodic surface. I argued that differences in brain function across individuals can be thus a primary source for variation in cultural symbolic system, and a primary factor to distinguish faithful transmitters from innovators. To our knowledge, this study is the first demonstration that social learning, transmission and cultural adaptation are constrained and driven by individual differences in the functional organization of sensory and memory systems. This finding opens new and intriguing questions. What would we expect by dividing populations of individuals by auditory processing capabilities (bad-learners versus good-learners)? How would the different processing capacities affect the cultural transmission of complex information?

In the last empirical chapter (**chapter 4**) I replicated and extended previous behavioral and neurophysiological findings on the temporal domain of music. I did this in two independent studies. In the first study, I used MGSGs as laboratory model of cultural evolution of symbolic system with signals (here equitone temporal patterns) associated to meanings (simple and compound emotions). I observed how different properties of the symbolic code emerged and changed across generations. I found near-periodic rhythms arising in the signaling systems, with other properties (i.e., syncopation, compression, and compositionality) remaining relatively unchanged. I speculated that temporal regularity, a universal feature of musical languages, might partially result from constraints on temporal processing and transmission processes. In the second study I combined signaling games and electrophysiology (as in **chapter 2**). Here too, individuals with the worse information processing capabilities (i.e., longer MMN) were found to introduce more regular, isochronic rhythms. All these findings (**chapters 3** and **4**) provide the first neurophysiological evidence in cultural transmission research that the tighter the ‘information processing bottleneck’ is, the greater the pressures for regularity and compression. Moreover, they suggest that individual small biases for temporal or melodic regularities, as revealed by neural predictors, amplify during cultural transmission through different mechanisms of cultural selection, to produce universal patterns of musical systems. We show that simple laboratory models of cultural transmission, and canonical paradigms of electrophysiology, can in principle address these complex issues.

5.2. Cultural evolution of music regularities in the laboratory

As I reviewed in **chapter 1**, two main theoretical lines in the field of music evolution tried to account in the past decades for why music is the way it is. On one hand, theories stressing the cultural nature of music; on the other hand, theories stressing its biological nature. In the same chapter, I argued that an alternative, more plausible solution to explain how music came to be as it is now, may reside in the middle of these two accounts: music as a socio-cultural artifact, but also adapted to cognitive pressures arising from cultural transmission (Merker, 2006; Merker et al., 2015).

In this thesis I was the first to *systematically* address this hypothesis in the laboratory, for two basic aspects of musical structure, namely rhythm and melody. Our findings (**chapters 2** and **4)** seem to support this view. I showed that when artificial tone sequences are culturally transmitted along diffusion chains of several individuals, they progressively develop regularities typical of their domain, in particular pitch proximity and continuity, melodic symmetry, and temporal isochronicity. Notably, these features are found in most of musical cultures of the world. The pervasive use, across several world cultures, of small pitch intervals organized in ascending or descending melodic directions, and equally spaced in time, was noticed by several scholars during the last century (e.g., Dowling, 1967; Brown & Jordania, 2013; Savage et al., 2015; Harwood, 1976), and so the use of melodic mirror forms (Dowling & Angeles, 1972; Tovey, 1956). The presence of these regularities in music has been often related to universals principles of pattern perception (Huron, 2001; Trehub 2000; Terhardt, 1987), and possibly explained by the recruitment of these principles during cultural transmission (Trehub, 2015; Merker 2006). To this day, only one study has tested this idea in the laboratory, and only in the domain of musical rhythm (Ravignani et al., 2016). I argued that basic principles of perceptual organization (Deutsch, 2012) and cognitive economy (Tamariz & Kirby, 2015), brought out through cultural transmission, may in principle explain these results.

Up to now, the use of classification methods on massive music databases including material of the past centuries, music notation or audio recordings, has been the main source of knowledge to provide insights into both the diversity and the unity of music (Lomax, 1987). In these works, music of the world was classified according to general phenetic (surface) similarities. The digitization of these databases, allowed researchers to use more sophisticated and rigorous statistical methods, including the phylogenetic methods, which explore the possible evolutionary mechanisms at the origin of these similarities (Savage & Brown, 2007). The recent use of other advanced statistical methods on large-scale music database (Serrà, 2013), was also useful to outline common historical trends on the evolution of musical styles, such as tracing changes in the frequencies of different musical traits in Western popular music (Mauch, MacCallum, Levy, & Leroi, 2015; Serrà, Corral, Boguñá, Haro, & Arcos, 2012), or to examine the perceptual mechanisms leading to the evolution of rhythmic and melodic features in Classical music (Zivic, Shifres, & Cecchi, 2013; Levitin et al., 2012). However, the unrepresentative sample of the data used (i.e., mostly Western music, and only of the past few centuries), with all the limitations inherent to statistical methods (Krumhansl, 2015), recommend some caution in the interpretation of these findings. The main weakness of these approaches is that, while they provide invaluable insights into real processes of cultural change, they do not allow manipulation of variables to test specific hypothesis on the mechanisms underlying the origins of these changes.

In this respect, I believe that historical and statistical methods can take advantage of hypothesis-driven experimental investigation. The conjunctive use of these methods can provide a better understanding of cultural phenomena (Mesoudi & Whiten, 2008), with the possibility to test hypotheses on cultural mechanisms only inferred with real-world data. Our experimental paradigm was effective in modelling both music structural regularities (on both rhythmic and melodic dimensions), and characteristics of cultural transmission and evolution, in particular of oral traditions (see section 1.7.1 for a comparison of MGSGs with iterated learning models). The transmission of musical cultures has been accomplished for hundreds (maybe thousands) of years by oral/aural methods, such as listening, repeating, and memorizing. Our model captures these aspects. The bottleneck introduced by human learning and memory constraints, transmission after transmission, may have produced some of the musical structures observed in the world, such as proximity and isochronicity, and may also explain why the contour (melodic and rhythmical) seems to be the musical dimension more preserved and characteristic of traditional folk music (Poladian, 1951). Notably, the social pressures introduced in our model made unnecessary any artificial manipulation for that to happen (Kirby et al., 2008; Verhoef, 2012). With this in mind, I suggest that the MGSGs can be used as a potential experimental methodology to test hypotheses on the cultural evolution of musical systems, a valuable alternative to the ILM.

5.3. Cognitive constraints and cultural selection on music transmission

So far, I have described limitations on memory and perception as the only source of change in music transmission. What about *selection* (Kleeman, 1985)? The role of ‘aesthetic selection’ has been examined in recent experimental research (MacCallum, Mauch, Burt, & Leroi, 2012). MacCallum et al. (2012) investigated whether Darwinian principles can account for how the preferences of listeners shape the evolution of music. The hypothesis here is that random transformative processes (*mutations* and *recombinations*) typically occur in the course of music transmission from one generation to another, producing variants that are eligible (or not) to become stable components of a musical repertoire. But while some Darwinian random processes, here cultural mutations, may in principle represent a ground for innovation in actual music (e.g., musical copy errors, random variations in performance etc.), they appear too rare to account for the rapid historical evolution of musical systems. In this respect, cognitive pressures may represent a major source of *directed*, non-random transformation of musical material, and thus, one major driving force, among others (e.g., cultural selection), in the evolution of music (Claidière, Kirby, & Sperber, 2012).

In support of this view, I showed (**chapters 2** and **4**) that inter-generational transmission is capable of deeply transforming the content of music-like material in non-random ways. I do not deny the just as important role played by selection on music evolution. I only argue that this role is different. If cognitive constraints are a major source of non-random innovation and transformation of musical variants, selection-like mechanisms are mainly responsible for the maintenance or changes of their frequencies over time (and/or within the population) (Mesoudi, 2015). When new variants are introduced in a musical repertoire, they can be maintained, die out, or spread, depending (often) on their intrinsic properties, especially aesthetic. The selection bias that makes consumers to favor one variant over another, according to its content, is known as content-dependent bias. This bias may be responsible of the difficulty of the twentieth century dodecaphonic music - in which all notes of the chromatic scales appear with the same frequency in a repeating row- to have a prominent role in the public scene (Ross, 2007). In other cases, selection can be context-biased: some variants are copied disproportionately more than others, or, on the contrary, they can disappear from the public scene, because of the political, social, and economical local context. The banishment of Mendelssohn music in Nazi Germany is one example of the effects of this selection mechanism (Haas, 2013). Finally, people can prefer some musical variants over others according to the popularity or the social prestige of the musician who performs or composes the music (or prestige biases). As a result of these selection mechanisms, music can become more appealing over time (MacCallum, 2012), or some of its variants, such as chord progressions in Western popular music, can be maintained unchanged for a long time (Serrà et al., 2012).

I acknowledge that music, more than language, is shaped by socio-intellectual and aesthetic factors, at least in the sense described above (Grout, 1980; Leroi, MacCallum, Mauch, & Burt, 2012; Temperley, 2004). In my thesis I did not addressed these factors. Rather, I decided to focus my investigation on cognitive constraints, designing our experiment with the specific intention to test their effects on cultural evolution (Boyd & Richerson, 1985; Caldwell et al., 2016). I argue that complex phenomena, such as musical transmission and evolution, can only be understood if abstracted and modelled in their more basic components (Richerson & Boyd, 1985).

Here, I focused on only one component: transmission (via interaction, alignment and coordination). All these pieces, selection and transmission included, need to be then integrated in a coherent theoretical framework if we want to provide a better understanding of how music evolved in the way it is now.

5.4. Symbolic systems adapting to the brain, cultural variation and the problem of linkage

One key tenet in cultural evolution research is the idea that the organization of cognition shapes culturally transmitted information over repeated episodes of transmission (Christiansen & Chater, 2001; Kirby, 2001; Deacon, 1997). The information acquired by an individual is typically transformed in non-random directions to fit functional aspects of his brain (Boyd & Richerson, 1985). The human brain/mind (or those of other species capable of cultural transmission) acts thus as a filter of information, so that more efficient forms outcompete others over time along several dimensions, among these conformity to processing constraints and memory capacities of the relevant neural networks (Merker, 2006). Literature commonly refers to this multi-dimensional filter as cognitive biases. In this thesis I preferred the use of another term, namely *cognitive constraints*. While constraints on information processing have been revealed by a wealth of cultural transmission studies (Fehér et al., 2009; Kashima, 2000; Kirby et al., 2008; Martin et al., 2014; Mesoudi & Whiten, 2004), their existence has been only assumed by observations (*i*) of behavioral data (*ii*) at individual level by iterated reproduction (e.g., Jacoby & McDermott, 2017), or (*iii*) by behavioral and simulation data at population-level, i.e., looking at arithmetical trends computed across several individuals or agents (e.g., Kirby et al., 2008; Ravignani et al., 2017; Kirby, 2001). It is quite surprising that up to now, no one tried to support the idea of cultural adaptation with neural data.

In this thesis I used both these approaches. I first showed that auditory symbolic systems adapt over time to the constraints imposed by the human brain, developing in the course of transmission prototypical melodic and rhythmic regularities. Then, I provided an initial evidence of the neural origins of this phenomenon, showing that information processing capabilities of individuals predict not only individual performance, but also the *direction* of cultural changes. I showed that individuals with less efficient auditory processing systems compress and regularize more than individuals with more efficient sensory systems. I speculated that each individual represents a different 'neural niche' (here the set of neural mechanisms and constraints underlying the processing and representation of tone sequences), through which cultural information must pass, and eventually adapt. Individual-level cultural transmission may thus provide a major source of variation in cultural symbolic systems.

How neural variation manifests and spread at population-level in more complex cultural dynamics is not easy to address. We may imagine a population of individuals with different biases in the acquisition and regularization of musical material. If the musical trait produced by ‘bad-learners’ (e.g., individuals with longer MMN latencies) is enough regular to ease its own acquisition, it may be selected (or adopted) by other bad learners, and would start to spread in the population. If bad-learners represent the majority of the population, traits would convergence to regular and easy-to-learn musical variants (Tamariz, Ellison, Barr, & Fay, 2014), with the complexity of musical systems (and the extent of variation) progressively decreasing over time. However, the outcome might be different with a different population structure, or when pressures for learnability were relieved by other vectors of transmission, such as music notation. In the first condition, we may consider a population where less constrained learners (individuals with shorter MMN latencies) represents the majority. These individuals would not suffer of pressures for regularity, and may favor the second because the more frequent. In this condition complexity could persist over time. Music notation may also counterbalance pressures for simplicity. In one experiment, Tamariz and Kirby (2015) compared two conditions in transmission chains, one where the cultural material (drawings) had to be transmitted across individuals by memory (‘remembering condition’) (Bartlett, 1932), and the other where it could be copied from view, without constraints on time (‘copy condition’). Only in the first condition, when participants were given a limited time to observe the drawings, the system decreased in complexity, becoming graphically simpler and schematic (i.e., resembling commonly known semantic categories). In the copy condition, barring few changes, complexity of the drawings remained unchanged. These findings suggest that, whenever memory and perception are relieved by their limits and constraints, using other channels of transmission (such as notation), the cultural system is less sensitive to pressures for regularities, and can maintains its original complexity. In this regard, teaching and written notation are two fundamental ingredients of cumulative cultural evolution, or ‘ratchet effect’ (Tomasello, 2009; Caldwell, Atkinson, & Renner, 2016; Caldwell & Millen, 2008), and can explain the great achievements of Western classic tonal music (Blanning, 2008). These predictions could be tested in the laboratory using microsocieties. We could create a condition with an high proportion of constrained learners and a low proportion of less constrained learners (and vice-versa for the second condition), and observe how variants emerge and spread in that populations using signaling games as a model of transmission and interaction. Would we observe the dynamics predicted? For more complex population dynamics, with innovators, copiers, bad-copiers, and egocentric individuals (unwilling to copy variants of the others), the effects are less predictable. Theoretical and computational models would be here necessary to lay some foundations, which then should be tested in the laboratory with real human learners.

The index of information processing used in this thesis, the MMN, is an early component in the information processing stages. In addition, it represents only the ‘tip of the iceberg’: it reflects constraints on transformational mechanisms, which are supposed to be upstream. These mechanisms operate on auditory sensory memory and include perceptual categorization, where continuous acoustic information is converted to discrete auditory events, and mechanisms of groupings, where auditory events are segmented based on different principles, i.e., proximity, similarity, and continuity (Deutsch, 2012). Operations in auditory memory, such as a regularity violation of an incoming sound patterns, are then (partly) reflected by the mismatch-negativity. These two mechanisms operate in the first 100 ms from the onset of the last heard events (Snyder, 2000). This leads us to speculate on the low-level (pre-attentive) nature of some of the cognitive constraints at work in the cultural transmission of symbolic systems, musical systems in particular.

I will conclude with a final point that concerns the problem of linkage (already discussed in **chapter 1**). Up to now, this problem has been framed at only two levels of explanation. The first, or *code level*, links code changes introduced by single individuals to population-level code patterns (Cavalli-Sforza & Feldman, 1981). The second, or *psychological level*, links (invariant) individual-level psychological constraints to population-level code patterns (Kirby et al., 2008). In this thesis I moved one step deeper, introducing the idea of the *neural level*. I showed that it is in principle possible to build experiments and to obtain data that link individual-level neural information processing with population-pattern phenomena, in the simplest case between two individuals in a coordination game. Moreover, in comparison to the psychological level, here I introduce the concept of *variation*. In some sense, this idea generates many questions to address. Which outcome would we expect by implementing differential constraints in a population of computational agents? Do we expect populations of good learners to reorganize a cultural system differently than a population of bad learners? To what extent teaching or written notation could counterbalance this result? Is it possible to find evidence for a further, *structural* level, where differences in brain structure would manifest as cultural variants? And many more. I leave these fascinating issues for future investigation.

5.5. Conclusions

In this thesis, I have investigated the cultural evolution of musical structure in the laboratory, by combining methods and paradigms employed in economics, evolutionary linguistics and cognitive neuroscience. I have argued for the importance of cognitive constraints in the emergence of both universal and peculiar aspects of human symbolic systems, music in particular. Perhaps more interestingly, I have provided an initial support for the neural origins of cultural changes, that need however to be further investigated. Last, I hope to have shown that aspects of music change can be studied experimentally in the laboratory, and that this research can be as fruitful as it is for language. If we want to undertake this task, we would do well to follow the concepts, methods, and tools of evolutionary linguistics, a discipline which has been through similar issues over the past decades (Ravignani & Gingras, 2013).

References

A

Aaltonen, O., Eerola, O., Lang, A. H., Uusipaikka, E., & Tuomainen, J. (1994). Automatic discrimination of phonetically relevant and irrelevant vowel parameters as reflected by mismatch negativity. *The Journal of the Acoustical Society of America*, *96*(3), 1489–1493. doi:10.1121/1.410291

Adams, C. R. (1976). Melodic contour typology. *Ethnomusicology*, *20*, 179-215. doi:10.2307/851015

Alain, C., Achim, A., & Woods, D. L. (1999). Separate memory-related processing for auditory frequency and patterns. *Psychophysiology*, *36*(6), 737–744. doi:10.1111/1469-8986.3660737

Alain, C., Cortese, F., & Picton, T. W. (1999). Event-related brain activity associated with auditory pattern processing. *Neuroreport*, *10*(11), 2429–2434. doi:10.1097/00001756-199810260-00037

Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., & Näätänen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology*, *77*(2), 151–155. doi:10.1016/0168-5597(90)90031-8

Arbib, M. A. (2013). *Language, Music, and the Brain: A Mysterious Relationship*. Cambridge, MA: MIT Press.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning & Memory* , *9*(3), 138–150.

B

Baggio, G. (2012). Selective alignment of brain responses by task demands during semantic processing. *Neuropsychologia*, *50*(5), 655–665. doi:10.1016/j.neuropsychologia.2012.01.002

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, *26*(9), 1338–1367.

Baldeweg, T., Richardson, A., Watkins, S., Foale, C., & Gruzelier, J. (1999). Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. *Annals of Neurology*, *45*(4), 495–503.

Bartel, C. (2006). Musical Thought and Compositionality. *Postgraduate Journal of Aesthetics*, *3*(1), 25–36.

Bartlett, F. C. (1932). *Remembering: An Experimental and Social Study*. Cambridge, UK: Cambridge University Press.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.99875-4.

Bharucha, J., Curtis, M., Paroo, K. (2011). Musical communication as alignment of brain states. In *Language and Music as Cognitive Systems*. Oxford: Oxford University Press.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100–130. doi:10.1016/j.cognition.2005.11.007

Bishop, D. V. M. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: where are we, and where should we be going? *Psychological Bulletin*, *133*(4), 651–672. doi:10.1037/0033-2909.133.4.651

Blacking, J., Byron, R., & Nettl, B. (1995). *Music, Culture, and Experience: Selected Papers of John Blacking*. Chicago, IL: University of Chicago Press.

Blanning, T. C. (2008). *The triumph of music: The rise of composers, musicians and their art.* Harvard: Harvard University Press.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Science of the United States of America*, 98, 11818– 11823.

Boh, B., Herholz, S. C., Lappe, C., & Pantev, C. (2011). Processing of complex auditory patterns in musicians and nonmusicians. *PloS One*, *6*(7), e21458. doi:10.1371/journal.pone.0021458

Boyd, R., & Richerson, P. J. (1988). *Culture and the Evolutionary Process*. Chicago, IL: University of Chicago Press.

Boyd, R., & Richerson, P. J. (2005). *The Origin and Evolution of Cultures*. Oxford, UK: Oxford University Press.

Bottcher□Gandor, C., & Ullsperger, P. (1992). Mismatch Negativity in Event□Related Potentials to Auditory Stimuli as a Function of Varying Interstimulus Interval. *Psychophysiology, 29*(5), 546-550.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Bregman, A. S., & Ahad, P. A. (1995). Demonstrations of auditory scene analysis: The perceptual organization of sound. *The Journal of the Acoustical Society of America*, *95*(2), 1177. doi:10.1016/s0166-2236(96)01022-3

Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, *2*(3), 177–226. doi:10.1016/j.plrev.2005.06.001

Brown, S., & Jordania, J. (2013). Universals in the world’s musics. *Psychology of Music*, 41(*2*), 229-248. doi:10.1177/0305735611425896

C

Caldwell, C. A., & Millen, A. E. (2008). Review. Studying cumulative cultural evolution in the laboratory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1509), 3529–3539. doi:10.1098/rstb.2008.0133

Caldwell, C. A., & Smith, K. (2012). Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PloS One*, *7*(8), e43807. doi:10.1371/journal.pone.0043807

Caldwell, C. A., Atkinson, M., & Renner, E. (2016). Experimental approaches to studying cumulative cultural evolution. *Current directions in psychological science*, 25(3), 191-195. doi: 10.1177/0963721416641049

Caldwell, C. A., Cornish, H., & Kandler, A. (2016). Identifying innovation in laboratory studies of cultural evolution: rates of retention and measures of adaptation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1690), 2015013. doi:10.1098/rstb.2015.0193

Cardenas, V. A., Chao, L. L., Blumenfeld, R., Song, E., Meyerhoff, D. J., Weiner, M. W., & Studholme, C. (2005). Using automated morphometry to detect associations between ERP latency and structural brain MRI in normal adults. *Human Brain Mapping*, *25*(3), 317–327. doi:10.1002/hbm.20103

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). Cultural transmission and evolution: a quantitative approach. *Monographs in Population Biology*, *16*, 1–388. doi:10.1016/0160-9327(82)90036-9

Chater, N., & Vitányi, P. (2003). Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*(1), 19–22. doi:10.1016/s1364-6613(02)00005-0

Chomsky, N. (1980). Rules and representations. *The Behavioral and Brain Sciences*. *3*(1), 1.

Choudhury, N., & Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: prediction to 3 and 4 year language and cognitive abilities. *Clinical Neurophysiology*, *122*(2), 320–338. doi:10.1016/j.clinph.2010.05.035

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *The Behavioral and Brain Sciences*, *31*(05), 489–489. doi:10.1017/s0140525x08004998

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *The Behavioral and Brain Sciences*, *39*, 1-52. doi:10.1017/s0140525x1500031x

Christiansen, M. H., & Kirby, S. (2003). *Language Evolution*. Oxford, UK: Oxford University Press.

Claidière, N., Kirby, S., & Sperber, D. (2012). Effect of psychological bias separates cultural from biological evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(51), E3526. doi:10.1073/pnas.1213320109

Cohen, R. A. (2014). *The Neuropsychology of Attention* (2nd ed.). New York, NY: Plenum Press.

Conard, N. J., Malina, M., & Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature*, *460*(7256), 737–740. doi:10.1038/nature08169

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*, 340–345. New York, NY: Curran Associates.

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. doi:10.1111/j.1467-9922.2009.00540.x

Cross, I. (2005). Music and meaning, ambiguity and evolution. In D. Miell, R. MacDonald, & D. Hargreaves (Eds.), *Musical Communication* (pp. 27-43). Oxford, UK: Oxford University Press.

Cross, I., & Woodruff, G. E. (2009). Music as a communicative medium. In R. Botha, & C. Knight (Eds.), *The Prehistory of Language*, (pp. 77–98). New York, NY: Oxford University Press.

Cross, I., Fitch, T., Aboitiz, F., Iriki, A., Jarvis, E., Lewis, J., Liebal, K., Merker, B., Stout, D., Trehub, S. E. (2013). Culture and evolution. In M.A. Arbib (Ed.), *Language, music, and the brain: a mysterious relationship* (pp. 541-562). Cambridge, MA: MIT Press.

D

Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems*, 153, 1150017.

Darwin, C. (1888). *The descent of man, and selection in relation to sex*. London, UK: John Murray.

Davies, S. (1994). *Musical Meaning and Expression*. Ithaca, NY: Cornell University Press.

de Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, *28*(4), 441–465.
Deacon, T. (1997). *The symbolic species*. New York, NY: Nortonµ.

Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences*, 104(*26*), 10944-10949. doi: 10.1073/pnas.0610848104

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*(2), 384–398.

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews. Neuroscience*, *16*(4), 234–244. doi:10.1038/nrn3924

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Deutsch, D. (2012). The processing of pitch combinations (3rd ed.). In D. Deutsch (Ed.), *The Psychology of Music* (pp. 249–325). New York, NY: Academy Press.

Dienes, Z., Kuhn, G., Guo, X. Y., & Jones, C. (2012). Communicating structure, affect and movement: Commentary on Bharucha, Curtis & Paroo. In Rebuschat, P., Rohrmeier, M., Cross, I. Hawkins (Eds.), *Language and Music as Cognitive Systems*. Oxford: Oxford University Press.

Dowling, W. J. (1967). *Rhythmic fission and the perceptual organization of tone sequences*. Unpublished doctoral dissertation. Cambridge, MA: Harvard University Press

Dowling, W. J. (1971). Recognition of inversions of melodies and melodic contours. *Perception &*

Psychophysics, *9*(3), 348–349. doi:10.3758/bf03212663

Dowling, W. J. (1972). Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics*, 12(5), 417–421. doi:10.3758/bf03205852

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

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. Orlando, FL:Academic Press.

Drake, C., & Bertrand, D. (2003). The Quest for Universals in Temporal Processing in Music. *The Cognitive Neuroscience of Music*, 21–31. doi:10.1093/acprof:oso/9780198525202.003.0002

Drake, C., & Botte, M. C. (1993). Tempo sensitivity in auditory sequences: evidence for a multiple-look model. *Perception & Psychophysics*, *54*(3), 277–286. doi:10.3758/bf03205262

Du, S., Tao, Y., & Martinez, A. M. (2014). Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(15), E1454–E1462. doi:10.1073/pnas.1322355111

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473(7345), 79-82.

E

Ekman, P., & Friesen, W. V. (2003). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Los Altos, CA: Malor Books.

Esper, E. A. (1925). A technique for the experiment investigation of associative interference in artificial linguistic material. *Language Monographs*.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *The Behavioral and Brain Sciences*, *32*(5), 429–448.

F

Fay, N., & Ellison, T. M. (2013). The cultural evolution of human communication systems in different sized populations: usability trumps learnability. *PloS one, 8*(8), e71781.

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The Interactive Evolution of Human Communication Systems. *Cognitive Science*, 34(*3*), 351-386.

Fehér, O., Wang, H., Saar, S., Mitra, P. P., & Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature*, *459*(7246), 564–568. doi:10.1038/nature07994

Feld, S., & Keil, C. (1994). *Music Grooves: Essays and dialogues.* Chicago: University and Chicago Press.

Fitch, W. T. (2005). The evolution of music in comparative perspective. *Annals of the New York Academy of Sciences*, *1060*(1), 29–49. doi:10.1196/annals.1360.004

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition*, *100*(1), 173–215. doi:10.1016/j.cognition.2005.11.009

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge, UK: Cambridge University Press.

Fitch, W. T. (2011). Unity and diversity in human language. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1563), 376–388. doi:10.1098/rstb.2010.0223

Fitch, W. T. (2012). The biology and evolution of rhythm: Unraveling a paradox. *Language and Music as Cognitive Systems*, 73–95. doi:10.1093/acprof:oso/9780199553426.003.0009

Fitch, W. T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1664), 20140091. doi:10.1098/rstb.2014.0091

Fitch, W. T. (2017). Cultural evolution: Lab-cultured musical universals. *Nature Human Behaviour*, 1, 0018.

Flynn, E., & Whiten, A. (2008). Cultural Transmission of Tool Use in Young Children: A Diffusion Chain Study. *Social Development* , *17*(3), 699–718. doi:10.1111/j.1467-9507.2007.00453.x

Ford, J. M., & Hillyard, S. A. (1981). Event-Related Potentials (ERPs) to Interruptions of a Steady Rhythm. *Psychophysiology*, *18*(3), 322–330. doi:10.1111/j.1469-8986.1981.tb03043.x

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–236. doi: 10.3758/s13421-014-0461-7

Fraisse, P. (1946). Contribution a l'étude du rythme en tant que forme temporelle. *Journal de psychologie normale et pathologique*, *39*, 283-304.

Fraisse, P. (1956). *Les structures rhythmiques.* Louvain, Belgium: Publications Universitaires de France.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 203-254). New York, NY: Academy Press.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience*, *16*(6), 1010–1021.

G

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, 29(*5*), 737-767. doi:10.1207/s15516709cog0000_34

Galantucci, B. (2009). Experimental semiotics: a new approach for studying communication as a form of joint action. *Topics in cognitive science,* 1(*2*), 393-410. doi: 10.1111/j.1756-8765.2009.01027.x.

Garrod, S., Fay, N., Rogers, S., Walker, B., & Swoboda, N. (2010). Can iterated learning explain the emergence of graphical symbols? *Interaction Studies*, *11*(1), 33–50. doi:10.1075/bct.45.03gar

Geiser, E., Ziegler, E., Jancke, L., & Meyer, M. (2009). Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex*, *45*(1), 93–102.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186. doi:10.1016/s1364-6613(00)01467-4

Gordon, R. L., Key, A., & 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. doi:10.1093/scan/nst017

Gottselig, J. M., Brandeis, D., Hofer-Tinguely, G., Borbély, A. A., & Achermann, P. (2004). Human central auditory plasticity associated with tone sequence learning. *Learning & Memory* , *11*(2), 162–171. doi:10.1101/lm.63304

Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426(*6965*), 435-439.

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.

Grout, D. J. (1980). *A history of Western music* (3rd ed.). New York, NY: WW Norton & Company, Inc.

H

Haas, M. (2013). *Forbidden Music: The Jewish Composers Banned by the Nazis.* New Haven: Yale University Press.

Hamming, R. W. (1986). *Coding and Information Theory.* Upper Saddle River, NJ: Prentice-Hall, Inc.

Hannagan, T., Amedi, A., Cohen, L., Dehaene-Lambertz, G., & Dehaene, S. (2015). Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. *Trends in Cognitive Sciences*, *19*(7), 374–382.

Harwood, D. L. (1976). Universals in music: A perspective from cognitive psychology. *Ethnomusicology*, 521-533.

Hockett, C. F. (1960). The origin of speech. *Scientific American, 203,* 89–96.

Honing, H., ten Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1664), 20140088. doi:10.1098/rstb.2014.0088

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.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language learning and development,* 1(*2*), 151-195.

Hurford, J. (1999). The evolution of language and languages. In R. Dunbar, C. Knight, & C. Power (Eds.), *The evolution of culture* (pp. 173-193). Edinburgh, UK: Edinburgh University Press.

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

Huron, D. (2001). Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles. *Music Perception*, *19*(1), 1–64. doi:10.1525/mp.2001.19.1.1

Huron, D. (2012). Two challenges in cognitive musicology. *Topics in Cognitive Science*, *4*(4), 678–684. doi:10.1111/j.1756-8765.2012.01224.x

Huron, D. B. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

J

Jackendoff, R. (2009). Parallels and non parallels between Language and Music. *Music Perception*, *26*(3), 195–204. doi:10.1525/mp.2009.26.3.195

Jackendoff, R. S., & Lerdahl, F. (1983). *A generative theory of tonal music*. Cambridge, MA.: MIT Press.

Jacoby N., & McDermott J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology.* In press corrected proof.

Juslin, P. N., & Sloboda, J. A. (2001). *Music and emotion: Theory and research*. Oxford, UK: Oxford University Press.

Justus, T., & Hutsler, J. J. (2005). Assessing Innateness and Domain Specificity. *Music Perception*, *23*(1), 1–27. doi:10.1525/mp.2005.23.1.1

K

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, 14(*2*), 288-294.

Kashima, Y. (2000). Maintaining Cultural Stereotypes in the Serial Reproduction of Narratives. *Personality & Social Psychology Bulletin*, *26*(5), 594–604. doi:10.1177/0146167200267007

Kempe, V., Gauvrit, N., & Forsyth, D. (2015). Structure emerges faster during cultural transmission in children than in adults. *Cognition*, *136*, 247–254. doi:10.1016/j.cognition.2014.11.038

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

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.

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 Laurea (Eds.), *Simulating the Evolution of Language* (pp. 121–147). New York, NY: Springer Verlag.

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 United States of America*, *105*(31), 10681–10686.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(12), 5241–5245.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology, 28,* 108–114. doi:10.1016/j.conb.2014.07.014

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102. doi:10.1016/j.cognition.2015.03.016

Kivy, P. (1980). *The corded shell: Reflections on musical expression*. Princeton, NJ: Princeton University Press.

Kleeman, J. E. (1985). The parameters of musical transmission. *Journal of Musicology*, 1–22. doi:10.2307/763720

Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, *8*(2), 89–105. doi:10.1016/j.plrev.2011.04.004

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences, 9*(2), 578-584. doi:10.1016/j.tics.2005.10.001

Koopman, C., & Davies, S. (2001). Musical Meaning in a Broader Perspective. *The Journal of Aesthetics and Art Criticism, 59*(3), 261–273. doi:10.1093/acprof:oso/9780199608775.003.0007

Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., & Koch, D. B. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science, 273*(5277), 971–973. doi:10.1126/science.273.5277.971

Krumhansl, C. L. (2015). Statistics, Structure, and Style in Music. *Music Perception, 33*(1), 20–31.

Kujala, T., Kallio, J., Tervaniemi, M., & Näätänen, R. (2001). The mismatch negativity as an index of temporal processing in audition. *Clinical Neurophysiology*, *112*(9), 1712–1719.

L

Ladinig, O., Honing, H., Hááden, G., & Winkler, I. (2009). Probing attentive and preattentive emergent meter in adult listeners without extensive music training. *Music Perception, 26*(4), 377–386.

Lang, A. H., Eerola, O., Korpilahti, P., Holopainen, I., Salo, S., & Aaltonen, O. (1995). Practical issues in the clinical application of mismatch negativity. *Ear and Hearing, 16*(1), 118–130.

Large, E. W. (2010). Neurodynamics of Music. In M. R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Music Perception* (pp. 201–231). New York, NY: Springer.

Leman, M. (1992). The theory of tone semantics: Concept, foundation, and application. *Minds and Machines, 2*(4), 345–363. doi:10.1007/bf00419418

Lerdahl, F. (1992). Cognitive constraints on compositional systems. *Contemporary Music Review*, *6*(2), 97–121. doi:10.1093/acprof:oso/9780198508465.003.0010

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

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, 109*(51), E3527–E3527. doi:10.1073/pnas.1214445109

Levitin, D. J., Chordia, P., & Menon, V. (2012). Musical rhythm spectra from Bach to Joplin obey a 1/f power law. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(10), 3716–3720.

Lewis, D. (1969). *Convention: A Philosophical Study.* Cambridge, MA: Harvard University Press

Lewontin, R. (2000). Does culture evolve? In M. Aloisi, B. Battaglia, E. Carafoli, & G.A. Danieli (Eds.), *The origin of humankind.* Amsterdam, the Netherlands: IOS Press.

Lomax, A. (1977). Universals in song. *World of Music, 19,* 117-129.

Lomax, A. (1980). Factors of musical style. *Theory & Practice: Essays Presented to Gene Weltfish*.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213. doi:10.3389/fnhum.2014.00213

Loui, P., Wessel, D., & Kam, C. H. (2006). Acquiring new musical grammars: a statistical learning approach. In R. sun, & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 1711–1716). Vancouver, BC: Curran Associates, Inc.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (3rd ed.). Cambridge, MA: MIT Press.

Lumaca, M., & Baggio, G. (2016). Brain potentials predict learning, transmission and modification of an artificial symbolic system. *Social Cognitive and Affective Neuroscience,* nsw112.

M

MacCallum, R. M., Mauch, M., Burt, A., & Leroi, A. M. (2012). Evolution of music by public choice. *Proceedings of the National Academy of Sciences of the United States of America, 109*(30), 12081–12086.

Martin, D., Hutchison, J., Slessor, G., Urquhart, J., Cunningham, S. J., & Smith, K. (2014). The spontaneous formation of stereotypes via cumulative cultural evolution. *Psychological Science, 25*(9), 1777–1786. doi:10.1177/0956797614541129

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.

Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science, 2*(5), 150081. doi:10.1098/rsos.150081

McDermott, J., & Hauser, M. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception, 23*(1), 29–59. doi:10.1525/mp.2005.23.1.29

McNamara, L., & Ballard, M. E. (1999). Resting arousal, sensation seeking, and music preference. *Genetic, Social, and General Psychology Monographs*, 125, 229–250.

Meyer, L. B. (1956). *Emotion and meaning in music.* Chicago: University of Chicago Press.

Meyer, L. B. (1957). Meaning in music and information theory. *The Journal of Aesthetics and Art Criticism,* 15, 412-424.

Menning, H., Roberts, L. E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *Neuroreport, 11*(4), 817–822. doi:10.1097/00001756-200003200-00032

Merker, B. (2002). Music: the missing Humboldt system. *Musicae Scientiae,* 6(*1*), 3-21.

Merker, B. (2006). The Uneven Interface Between Culture and Biology in Human Music. *Music Perception*, *24*(1), 95-98. doi:10.1525/mp.2006.24.1.95

Merker, B., Morley, I., & Zuidema, W. (2015). Five fundamental constraints on theories of the origins of music. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1664), 20140095. doi:10.1098/rstb.2014.0095

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

Mesoudi, A. (2007). Using the methods of experimental social psychology to study cultural evolution. *Journal of Social, Evolutionary & Cultural Psychology, 1*(2), 35. doi:10.1037/h0099359

Mesoudi, A. (2015). Cultural Evolution: A Review of Theory, Findings and Controversies. *Evolutionary Biology*, 1–17. doi:10.1007/s11692-015-9320-0

Mesoudi, A., & Whiten, A. (2004). The Hierarchical Transformation of Event Knowledge in Human Cultural Transmission. *Journal of Cognition and Culture*, *4*(1), 1–24.

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. Series B, Biological Sciences*, *363*(1509), 3489–3501. doi:10.1098/rstb.2008.0129

Mesoudi, A., Whiten, A., & Laland, K. N. (2004). Perspective: is human cultural evolution Darwinian? Evidence reviewed from the perspective of the Origin of Species. *Evolution; International Journal of Organic Evolution*, *58*(1), 1–11. doi:10.1111/j.0014-3820.2004.tb01568.x

Miller, G., Wallin, N., Merker, B., & Brown, S. (2000). *The origins of music*. Cambridge, MA: MIT Press.

Miranda, E. R., Kirby, S., & Todd, P. (2003). On Computational Models of the Evolution of Music: From the Origins of Musical Taste to the Emergence of Grammars. *Contemporary Music Review*, *22*(3), 91–111.

Moore, J. K., & Guan, Y. L. (2001). Cytoarchitectural and axonal maturation in human auditory cortex. *Journal of the Association for Research in Otolaryngology*, *2*(4), 297–311. doi:10.1007/s101620010052

Moreno, M., & Baggio, G. (2015). Role asymmetry and code transmission in signaling games: an experimental and computational investigation. *Cognitive Science*, *39*(5), 918–943. doi:10.1111/cogs.12191

Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archaeology, and the Origins of Musicality*. Oxford, UK: Oxford University Press.

N

Näätänen, R., & Alho, K. (1997). Mismatch negativity-the measure for central sound representation accuracy. *Audiology and Neurotology*, *2*(5), 341–353. doi:10.1159/000259255

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*(6), 826–859. doi:10.1037//0033-2909.125.6.826

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. doi:10.1016/0001-6918(78)90006-9

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology*, *118*(12), 2544–2590.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *Neuroreport*, *4*(5), 503–506.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model.* Chicago: University of Chicago Press.

Nattiez, J. J. (1990). *Music and discourse: Toward a semiology of music.* Princeton University Press, 1990.

Nettl, B. (1983). *The Study of Ethnomusicology: Twenty-nine Issues and Concepts*. Chicago, IL: University of Illinois Press.

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 463–472). Cambridge, MA: MIT Press.

Novak, G., Ritter, W., & Vaughan, H. G. (1992). The Chronometry of Attention□Modulated Processing and Automatic Mismatch Detection. *Psychophysiology*. doi:10.1111/j.1469-8986.1992.tb01714.x

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. doi:10.1093/jole/lzw007

O

Orsini, A., Grossi, D., Capitani, E., Laiacona, M., Papagno, C., & Vallar, G. (1987). Verbal and spatial immediate memory span: normative data from 1355 adults and 1112 children. *Italian Journal of Neurological Sciences*, *8*(6), 539–548. doi:10.1007/bf02333660

P

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: a review. *International Journal of Psychophysiology*, *88*(2), 109–123. doi:10.1016/j.ijpsycho.2013.03.015

Pablos Martin, X., Deltenre, P., Hoonhorst, I., Markessis, E., Rossion, B., & Colin, C. (2007). Perceptual biases for rhythm: The Mismatch Negativity latency indexes the privileged status of binary vs non-binary interval ratios. *Clinical Neurophysiology*, *118*(12), 2709–2715.

Pagel, M., & Mace, R. (2004). The cultural wealth of nations. *Nature*, *428*(6980), 275–278.

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*, *110*(22), 8836–8841.

Patel, A. D. (2010). *Music, Language, and the Brain*. New York, NY: Oxford University Press.

Pekkonen, E., Rinne, T., & Näätänen, R. (1995). Variability and replicability of the mismatch negativity.

Electroencephalography and Clinical Neurophysiology, *96*(6), 546–554. doi:10.1016/0013-4694(95)00148-r

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691. doi:10.1038/nn1083

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.

Poladian, S. (1951). Melodic Contour in Traditional Music. *Journal of the International Folk Music Council*, 3, 30-35. doi:10.2307/835769

Port, R. F. (1991). Can complex temporal patterns be automatized? *The Behavioral and Brain Sciences*, *14*(04), 762–764. doi:10.1017/s0140525x00072447

Povel, D. J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology. Human Perception and Performance*, *7*(1), 3–18. doi:10.1037//0096-1523.7.1.3

R

Ravignani A., Delgado T. and Kirby S. (2016). The Cultural Evolution Of Structure In Music And Language. In S.G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.) *The Evolution of Language: Proceedings of the 11th International Conference*.

Ravignani, A., & Gingras, B. (2013). EvoMus: The evolution of music and language in a comparative perspective. In H. Lyn, & E.A. Cartmill (Eds.) *The Evolution of Language: Proceedings of the 10th International Conference*, Singapore: World Scientific.

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). Boston, MA: MIT Press.

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

Rodriguez Zivic, P. H., Shifres, F., & Cecchi, G. A. (2013). Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(24), 10034–10038. doi:10.1073/pnas.1222336110

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. doi:10.1371/journal.pone.0066174

Rohrmeier, M., Rebuschat, P., & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness and Cognition, 20*(2), 214–222. doi:10.1016/j.concog.2010.07.004

Ross, A. (2007). *The Rest is Noise: Listening to the Twentieth Century*. New York: Farrar, Straus and Giroux.

Rubin, D. C. (1997). *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes*. Oxford: Oxford University Press.

Rzeszutek, T., Savage, P. E., & Brown, S. (2012). The structure of cross-cultural musical diversity. *Proceedings. Biological Sciences/The Royal Society*, *279*(1733), 1606–1612.

S

Sable, J. J., Gratton, G., & Fabiani, M. (2003). Sound presentation rate is represented logarithmically in human cortex. *The European Journal of Neuroscience*, *17*(11), 2492–2496. doi:10.1046/j.1460-9568.2003.02690.x

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PloS ONE* 4(10): e7487. doi:10.1371/journal.pone.0007487

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *62*(6), 437–448.

Savage, P. E., & Brown, S. (2007). Toward a New Comparative Musicology. *Analytical Approaches to World Music*, 2(2), 148–197.

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(29), 8987–8992. doi:10.1073/pnas.1414495112

Sawa, G. D. (2002). Oral transmission in arabic music, past and present. *Oral Tradition*, *4*(1).

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

Schönpflug, U. (2008). *Cultural Transmission: Psychological, Developmental, Social, and Methodological Aspects*. Cambridge, UK: Cambridge University Press.

Schröger, E., Näätänen, R., & Paavilainen, P. (1992). Event-related potentials reveal how non-attended complex sound patterns are represented by the human brain. *Neuroscience Letters*, *146*(2), 183–186. doi:10.1016/0304-3940(92)90073-g

Schulze, H.-H. (1989). Categorical perception of rhythmic patterns. *Psychological Research*, *51*(1), 10–15. doi:10.1007/bf00309270

Schwartze, M., Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2011). Temporal regularity effects on pre-attentive and attentive processing of deviance. *Biological Psychology*, *87*(1), 146–151.

Scott-Phillips, T. C., & Kirby, S. (2010). Language evolution in the laboratory. *Trends in Cognitive Sciences*, *14*(9), 411–417. doi:10.1016/j.tics.2010.06.006

Scott-Phillips, T.C., Kirby, S., & Ritchie, G.R.S. (2009). Signalling signalhood and the emergence of communication. *Cognition*, 113(*2*), 226–233.

Sedlmeier, P., Weigelt, O., & Walther, E. (2011). Music is in the muscle: How embodied cognition may influence music preferences. *Music Perception: An Interdisciplinary Journal*, 28(*3*), 297-306.

Selten, R., & Warglien, M. (2007). The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(18), 7361–7366. doi:10.1073/pnas.0702077104

Seppänen, M., Brattico, E., & Tervaniemi, M. (2007). Practice strategies of musicians modulate neural processing and the learning of sound-patterns. *Neurobiology of Learning and Memory*, *87*(2), 236–247.

Serrà, J. (2013). What can we learn from massive music archives? In S. Müller, S.S. Narayanan, & Schuller B. (Eds.), *Dagstuhl seminar 13451: Computational Audio Analysis*. Wadern, Germany.

Serrà, J., Corral, A., Boguñá, M., Haro, M., & Arcos, J. L. (2012). Measuring the evolution of contemporary western popular music. *Scientific Reports*, *2*, 521. doi:10.1038/srep00521

Shannon, C. E. (1949). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*, 379-423.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of information*. University of Illinois Press, Urbana, Il.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York, NY: McGraw-Hill.

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: a characterization of the mind's bottleneck. *PLoS Biology*, *3*(2), e37. doi:10.1371/journal.pbio.0030037

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford, UK: Oxford University Press.

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford: Oxford University Press.

Smith, K., & Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1509), 3591–3603. doi:10.1098/rstb.2008.0145

Snyder, B. (2008). Memory for music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford Handbook of Music Psychology* (pp. 107–117). Oxford, UK: Oxford University Press.

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

Spike, M., Smith, K., & Kirby, S. (2016). Minimal pressures leading to duality of patterning. In S.G. Roberts, C. Cuskley, L. McCrohon, L. Barceló-Coblijn, O. Fehér, & T. Verhoef (Eds.) *The Evolution of Language: Proceedings of the 11th International Conference* (EVOLANG11), New Orleans.

Stevens, C., & Byron, T. (2009). Universals in music processing. In S. Hallam, I. Cross, & M. Thaut (Eds.) *Oxford handbook of music psychology* (pp. 14–23). Oxford, UK: Oxford University Press.

Sussman, E., Ritter, W., & Vaughan, H. G., Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*, *36*(1), 22–34.

Swarup, S., & Gasser, L. (2009). The Iterated Classification Game: A New Model of the Cultural Transmission of Language. *Adaptive Behavior*, *17*, 213-235. doi: 10.1177/1059712309105818

T

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

Tamariz, M. (2017). Experimental Studies on the Cultural Evolution of Language. *Annual Reviews of Linguistics*.

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

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, 8, 27-43

Tamariz, M., & Smith, A. D. (2008). Regularity in mappings between signals and meanings. *In Proceedings of the 7th International Conference on the Evolution of Language*, 315-322.

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.

Tamariz, M., Ellison, T. M., Barr, D. J., & Fay, N. (2014). Cultural selection drives the evolution of human communication systems. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1788), 20140488.

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.

Tavano, A., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schröger, E. (2014). Temporal regularity facilitates higher-order sensory predictions in fast auditory sequences. *The European Journal of Neuroscience*, *39*(2), 308–318. doi:10.1111/ejn.12404

Temperley, D. (2004). Communicative Pressure and the Evolution of Musical Styles. *Music Perception*, *21*(3), 313-337. doi:10.1525/mp.2004.21.3.313

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

Tervaniemi, M., Ilvonen, T., Karma, K., Alho, K., & Näätänen, R. (1997). The musical brain: brain waves reveal the neurophysiological basis of musicality in human subjects. *Neuroscience Letters*, *226*(1), 1–4.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J., & Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning & Memory* , *8*(5), 295–300. doi:10.1101/lm.39501

Tervaniemi, M., Schröger, E., & Näätänen, R. (1997). Pre-attentive processing of spectrally complex sounds with asynchronous onsets: an event-related potential study with human subjects. *Neuroscience Letters*, *227*(3), 197–200. doi:10.1016/s0304-3940(97)00346-7

Theisen-White, C. A., Oberlander, J., & Kirby, S. (2012). Systematicity and arbitrariness in novel communication systems. *Studies on the Emergence and Evolution of Human Communication*, 15–32. doi:10.1075/bct.45.02the

Tierney, A. T., Russo, F. A., & Patel, A. D. (2011). The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(37), 15510–15515.

Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, *372*(6501), 90-92. doi:10.1038/372090a0

Todd, J., Myers, R., Pirillo, R., & Drysdale, K. (2010). Neuropsychological correlates of auditory perceptual inference: a mismatch negativity (MMN) study. *Brain Research*, *1310*, 113–123. doi:10.1016/j.brainres.2009.11.019

Tolbert, E. (2001). Music and meaning: An evolutionary story. *Psychology of Music*, 29, 84-94.

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

Tovey, D. F. (1956). *The forms of music*. Cleveland: Meridian-Books.

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 of London. Series B, Biological Sciences*, *370*(1664), 20140089. doi:10.1098/rstb.2014.0089

Trainor, L., McFadden, M., Hodgson, L., Darragh, L., Barlow, J., Matsos, L., & Sonnadara, R. (2003). Changes in auditory cortex and the development of mismatch negativity between 2 and 6 months of age. *International Journal of Psychophysiology*, 51(1), 5–15.

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

Trehub, S. E. (2015). Cross-cultural convergence of musical features. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(29), 8809–8810.

Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Phil. Trans. R. Soc. B*, 370(1664), 20140096. doi:10.1098/rstb.2014.0096

Tremblay, K., Kraus, N., & McGee, T. (1998). The time course of auditory perceptual learning: neurophysiological changes during speech□sound training. *Neuroreport*, *9*(16), 3557.

V

van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2004). Grouping of sequential sounds—an event-related potential study comparing musicians and nonmusicians. *Journal of Cognitive Neuroscience*, *16*(2), 331–338. doi:10.1162/089892904322984607

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, *4*(4), 357–380. doi:10.1515/langcog-2012-0019

Verhoef, T., Kirby, S., & de Boer, B. (2013). Combinatorial structure and iconicity in artificial whistled languages. In M. Knauff (Ed.), *Proceedings of the 35th annual conference of the cognitive science societ*y (pp. 3669–3674). Berlin, Germany: Curran associates, Inc.

Von Hippel, P. (2000). Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals. *Music Perception*, *17*(3), 315–327. doi:10.2307/40285820

Vos, P. G., Mates, J., & van Kruysbergen, N. W. (1995). The perceptual centre of a stimulus as the cue for synchronization to a metronome: Evidence from asynchronies. *The Quarterly Journal of Experimental Psychology*, *48*(4), 1024–1040. doi:10.1080/14640749508401427

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music-brain responses to rhythmic incongruity. *Cortex*, *45*(1), 80–92. doi:10.1016/j.cortex.2008.05.014

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., & Østergaard, L. (2005). To musicians, the message is in the meter pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage*, *24*(2), 560–564.

W

Whiten, A., & Mesoudi, A. (2008). Establishing an experimental science of culture: animal social diffusion experiments. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 363*, 3489-3501. doi: 10.1098/rstb.2008.0134

Whiten, A., Spiteri, A., Horner, V., Bonnie, K. E., Lambeth, S. P., Schapiro, S. J., & de Waal, F. B. M. (2007). Transmission of multiple traditions within and between chimpanzee groups. *Current Biology*, *17*(12), 1038–1043. doi:10.1016/j.cub.2007.05.031

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem. *Memory* , *18*(6), 657–669. doi:10.1080/09658211.2010.501339

Winkler, I. (2007). Interpreting the Mismatch Negativity. *Journal of Psychophysiology*, 21(3-4), 147–163. doi:10.1027/0269-8803.21.34.147

Winkler, I., & Schröger, E. (1995). Neural representation for the temporal structure of sound patterns. *Neuroreport*, *6*(4), 690–694. doi:10.1097/00001756-199503000-00026

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(7), 2468–2471. doi:10.1073/pnas.0809035106

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.

X

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), 20123073–20123073. doi:10.1098/rspb.2012.3073

Z

Zatorre, R. J. (2013). Predispositions and plasticity in music and speech learning: neural correlates and implications. *Science*, *342*(6158), 585–589. doi:10.1126/science.1238414

Zatorre, R. J., Delhommeau, K., & Zarate, J. M. (2012). Modulation of auditory cortex response to pitch variation following training with microtonal melodies. *Frontiers in Psychology*, *3*, 544. doi:10.3389/fpsyg.2012.00544