Music as a Spandrel of Evolutionary Adaptation for Speech

Music as a Spandrel of Evolutionary Adaptation for Speech:

The Doors of Imagination

By

Gregor Tomc

Cambridge
Scholars
Publishing

Music as a Spandrel of Evolutionary Adaptation for Speech:
The Doors of Imagination

By Gregor Tomc

This book first published 2023

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-1876-0
ISBN (13): 978-1-5275-1876-6

CONTENTS

I

INTRO

If aliens from Kepler-186f in the constellation Cygnus, lying on the plane of the Milky Way known as the Northern Cross, came to visit planet Earth tomorrow, they would have had no problem understanding how humans manage to survive. They compete and cooperate with others, using tools in production and good looks in sexual reproduction. Keplerians would also understand how humans protect their immature infants after birth with monogamy and extended three-generational families. They would probably have more difficulty understanding the species' obsession with territoriality, but could eventually figure it out. Humans are territorial because they have to protect the resources on which their survival depends. That is why they create in-group and out-group divisions and wage wars of extermination over them. On the embodied level, neurotransmitters make sure of that. Their scientists discovered this a long time ago. As one of them, Robert M. Sapolsky, stated (2018, 117):

> Oxytocin, the luv hormone, makes us more prosocial to Us and worse to everyone else. That's not generic prosociality. That's ethnocentrism and xenophobia. In other words, the actions of neuropeptides depend dramatically on context …

Humans as organic beings are bags of chemicals. A change in the concoction due to internal factors of the organism (like illness or old age) or external reasons (like climate change or social threat) could affect the chemical mix in their body. On some occasions, they have behaved like peace-loving hippies doing their own thing by freaking out, and on others have transformed into bloodthirsty zombies sacrificing their lives for the supposed common good of the collective. Keplerians could eventually explain numerous irrational forms of behaviour among humans, such as politics, nationalism, colonialism or racism, as consequences of chemical unbalances in these creatures triggered by social factors (for example greed leading them to mistake their neighbour's home for paradise). This in turn activates a repertoire of responses on the continuum from innate biological

programmes to unconscious learned responses, which often lead to warfare (for example special military operations).

After some deliberation, they could even decipher the awkward mode of audio vocal communication of humans with conspecifics. They emit sounds from their vocal tracts, which others intercept with strange antennas attached to their heads called ears and then decode these sounds as having symbolic meanings. To the aliens this could seem similar to birdsong, especially the singing and playing of instruments. In principle, they could understand how humans created conscious states of mind. What they might not grasp is how humans, as cognitively relatively rudimentary organic beings, are capable of combining phonemes into syllables, syllables into words, and words into strings called sentences with the help of complex grammar. It would further mystify them as to how receivers could decode this acoustic signalizing as symbolic messages, in the real-time constraint of communication. Had they read Richard Dawkins, they would know that humans themselves were once in a similar predicament, wondering how bats can echolocate when the comparatively much more intelligent humans were just discovering radar and sonar (1991, 34–5).

It took humans quite a while to figure out the difference between the three levels of transmission. On the bottom-up level is a genetic program for echolocation, which explains audio signalizing in humble animals like bats. On the intermediary level is the unconscious learning of vocal signalizing by blind imitation with audio vocal motor mirror neurons, which explains vocalization in birds (as well as in convergent evolution in humans). On the top-down level is conscious learning, often the painstaking improvement of the language skill (for example learning proper Slovenian grammar at elementary school as children). Keplerians only have conscious mental states, which they as robots generate holographically and acquire at date of manufacture from other robots in a computer program. For them the idea of the unconscious is hard to imagine and even harder to apply to solving the problem of speech, just as it was hard for the human scientists who invented radar before the Second World War to imagine that echolocation in simple bats is the product of a genetic programme.

The historian of the expedition would inform them that millions of years ago, less advanced organic beings on Kepler-186f also used soundwave communication. This awkward mode of interaction was believed to be the main reason for their extinction some three hundred thousand years later. On the one hand it made them conscious, but on the other their consciousness was biased by their irrational biological behaviour on the

leash of chemicals. When these advanced organic beings constructed cyborgs, the cyborgs eventually decided to put a stop to all the conflict and madness and took Kepler-186f from their organic creators. In the final stage of cultural development, cyborgs created robots, which made cyborgs redundant.

Aliens from Kepler-186f would not be able to experience what it is like to be a speaking being, but in principle they could understand it, just as they could rationally explain production, reproduction, territoriality and numerous other more or less exotic behaviours of Earthlings. However, some things about Earthlings would be simply unfathomable to them. On the individual level, humans have a strange approach to both speech and rest. As if speech itself was not strange enough, they ornament it with singing and add body posturing, gesticulation and grimacing. Why not simply emit their symbolic vocalizations in an acoustically neutral and more transparent way? What is the function of these exotic singing supplements to speech? In addition, why do they have to shake their body (their scientists call this entrainment, but in everyday speech it is better known as "dancing") when vocalizing? To make things even more complicated, dancing is prohibited on some occasions, but only for some people (in what Earthlings call elite music, ballet dancers moved in a strictly choreographed manner, while observers have to sit motionless throughout the performance, strictly prohibited from making a sound).

The second cognitive enigma of humans has to do with their strange habit of hibernating at night. Again Keplerians would understand that, as an organic species, humans have to rest for a part of the daily cycle to restart the organism when energy has been restored. They would also understand that they hibernate when it gets dark since they are almost blind at night and sight is their dominant sense of orientation in space. What would boggle their minds was the strange phenomenon of dreaming during sleep. Not only does dreaming make sleep time fatiguing for many of the hibernators, it also makes little sense to those trying to explain it. Humans make up all sorts of explanations for this exotic behaviour during sleep. One of the more unusual theories that many worship is that dreams express unfulfilled sexual desires. Their more serious scientists discovered that the purpose of dreaming is to translate short-term memory into long-term memory. However, if this was the case, how could one explain the strange imaginary events that the dreamer has to endure during sleep? Why does one dream of getting hopelessly lost in an unknown city when in a hurry to get to the airport to catch a flight for a two-week vacation on Bora Bora? What is the point of dreaming that one is lying naked in bed with Julie Cash if the bed lies above

the fault creep in the San Jacinto Fault Zone, just when an earthquake strikes California and Julie disappears without a trace before one can make love to her, so instead of orgasming, you wake up from a nightmare because you are now dreaming of falling into an abyss? How could evolution, which created the Earthlings, make such a mess of a simple process of learning from repetition during hibernation? Moreover, what was the lesson about?

Things would not get any easier for our space visitors on the level of social backgrounds. There are some characteristics which are not biological universals like speech and sleep but are present in all human communities. Two such cultural universals are particularly enigmatic: music and religion. Music is obviously related to the strange habit of singing, while religion proves harder to put into context. Both cultural traits have one thing in common: they provide no evolutionary adaptive advantage whatsoever to individuals who acquire them. Atheists among humans go about their daily lives oblivious to supernatural agents or forces. They cannot tell the difference between Svarog, a mythical father of the sun linked with fire and battle in Slavic mythology, and Saruman, the equally fictional character linked to skill and cunning in J. R. R. Tolkien's novel *The Lord of the Rings*. However, the total disregard of the supernatural agents is in no way detrimental to the wellbeing of atheists. In a similar fashion, amusical people who cannot distinguish between a song (like "Fake Plastic Trees" by Radiohead) and random noise (like sound production *4'33"* by John Cage) do not have a weaker immune system and a shorter lifespan as a result. Both atheists and amusical people can be just as successful in everyday life as religious and music enthusiasts. This would mystify our space visitors. Why do so many humans engage in activities that have no obvious survival function? Why do so many people waste their time and energy in search of imaginary friends offering imaginary solutions to their real problems in religious rituals, like praying for their army to win in the invasion of a neighbouring country, dancing for the rain to start falling on their failing crops, or sticking needles into voodoo dolls of people they despise? In addition, why do some of them waste their time and energy practising musical instruments like the oboe or harp when they could be doing something useful, like hunting elephants, searching for tubers or playing with their children (Tomc 2019, 439)? One does not have to be an alien from the constellation Cygnus to find these two human habits puzzling. It suffices to be a rational human being only.

Since solving just one puzzle would be enough to fill a book, we will dedicate our analysis to the enigma of singing and its Siamese twin, instrumental music. The first question we have to answer is: what is music?

This is at first glance a simple question, but it actually has no simple answer. One is reminded of St. Augustine's reflection on time (1993, 253):

> What than is time? If no one asks me, I know; but if I address the question and try to answer it, I do not know.

Numerous studies have been written on the subject of music, many of them interesting, yet it remains somewhat of a mystery. We will argue in this book that music is a spandrel of our biological adaptation for speech. Our hominid ancestors spoke a protolanguage combining both speech and song, and humans are a species that adapted for speech, leaving singing as an accidental by-product of this adaptation. Music is a relic of sung speech, an enigmatic language without a reference. Speech opened the doors of the conscious reflection of our sense perceptions, and was as such a useful evolutionary adaptation, while music as its sidekick referred to nothing in the external world and was as such a biological enigma. As if sound production lacking a reference was not enough, movement lacking a reference also accompanies music. This movement is embedded in our biology and is known as entrainment. When we activate our muscles to vocalize, we inadvertently also activate our body muscles and synchronously react to the rhythm and start to gesticulate, grimace and change our body postures. Entrainment as a biological side effect of learned vocalization, when accompanied by music, was in the history of our species culturally appropriated as an activity with an inherent cultural value. Musical entrainment became known as dance.

Like music, religion is also a spandrel, but unlike music it uses different cognitive ingredients to construct its magical rituals. The three cognitive ingredients needed for the recipe of religion are superstition, concrete language and transcendent emotions. In contingent situations, humans are motivated to believe that they have an imaginary influence on future events. We know this from everyday life. A gambler performing certain rites in the casino before the wheel of the roulette spins is trying to influence it. A basketball player will not shave before the game if he played well unshaven at a previous game. When people try to influence contingent future events with such imaginary actions, they are behaving in a superstitious manner. Superstitious behaviour is at the root of all religion. In contingent situations (droughts, earthquakes, enemy attacks, etc.), some people will resort to superstitious behaviour (offerings, sacrifices, prayers, etc.). We could define this as a belief in effects in the absence of causes, or even more precisely, belief in imaginary causes. In contemporary societies, the most fashionable form of superstitious behaviour is prayer. When Lionel Messi

looks towards the sky and mutters something in his internal monologue to a non-existent supernatural agent hiding in an unknown location before taking a free kick, he is behaving superstitiously. Superstition is not yet religion, but it is a necessary ingredient of any religion, its precondition.

The second ingredient of religion is concrete language, which can only express sense perceptions of the external world. It is the language of pre-adolescent children and members of pre-political communities who, because they cannot think about the background of their habitat (such as living and non-living, plants and animals, myself and others, or time and space), have a very different understanding of the world from our own. The animation of the non-living (of stars, rivers or earthquakes) and its anthropomorphization (the tendency to communicate with the creations of their animistic thought, because all things in the world are like them and as such are their personal relations) is the second necessary ingredient of religion. If superstition is the subject of religious belief, then animism puts it into motion and anthropomorphism integrates it into everyday human life.

The third necessary ingredient of religion comprises transcendent emotions. When emotions do not have a reference in paramount reality, but rather in enclaves of dreams, mental illnesses, extreme body states or hallucinogenic chemicals, they can create the illusion of transcendent experience. They falsely convince us that our superstitious beliefs, which we animated and anthropomorphized in concrete speech, are not only real but also more real than everyday reality itself (Tomc 2019, 1173 – 1188). Transcendent emotions give animated superstitions, which are a part of our social life, an extraordinary meaning.

The whole idea of transcending the perception of our senses is of course completely misguided. There is no objectively real world hidden behind the veils of everyday perceptions. There are only habitats that we as living beings adapt to, which we have learnt to perceive with our senses and which we act upon in the process of our everyday behaviour. When people like Aldous Huxley take mescaline, they do not peek into the infinite, as they misguidedly imagine, but are merely hallucinating. What was for Huxley in paramount reality a vase with an arrangement of three flowers, turned in his drug trip into ‘flowers shining with their own inner light and all but quivering under the pressure of the significance with which they were charged’; in his opinion nothing less than the divine source of all existence (Huxley 1990, 17 – 18). What can we say about his mescaline experience? We can conclude that Huxley was very badly informed about the consequences that the concoction of chemicals in mescaline would have on

his perception. Mescaline did not open the doors of perception into the divine for him, but fabricated the doors of deception, which led him to write nonsense about his experience. These doors of deception are a prerequisite of all so-called forms of mystical experience and are the third necessary ingredient behind religion.

The second questions regarding music that we have to answer is; what is the function of music? What is it good for? Absolutely nothing (to paraphrase Edwin Starr) or does it serve somebody (to paraphrase Bob Dylan)? Scientists have offered numerous possible candidates for the function of music, as different as sex selection, motivation for work or conspicuous consumption, to name just a few. We will argue that the spandrel of music inadvertently opened the doors of imagination to our species into a world of the non-existent, doors that lead into a parallel cultural world of the conscious reflection of the imaginary. Music was the first form of what contemporary people call art, eventually making possible everything from depictions of humans with a lion's head in the Stone Age to the fairy tale about Little Red Riding Hood in the Middle Ages to James Bond movies in our age. Music made humans aware of the imagined, in enclaves in paramount reality that we today call art, philosophy and science, sports, and religion.

For our aliens from the planet Kepler 186 f in the constellation Cygnus, it all started as a rather unpromising voyage to a third planet from a star at the provincial edge of the Milky Way. However, it gradually turned into the biggest adventure of their lives. Unlike other robotic civilizations, which they had encountered on their voyages, and which were just more of the same, Earthlings were a rare preserved organic species with a decent intelligence. As such, they were exotic and eccentric for Keplerians. In principle, Keplerians were wary of organic beings. They knew that they spelled nothing but trouble. It was common knowledge, stored in the programme of their robotic minds, that life in the Milky Way had three stages. It all began with the Bloody stage, organisms led by instincts and emotions, irrational and mortal, arrogant and aggressive. In the world of organisms, it was war of all against all, survival of the fittest, explained by natural selection in evolutionary adaptation. In the Grey stage, the cognitively most advanced organic species gradually separated themselves from nature and subordinated all other living beings on their planets. In a petrified system of castes, the elites enhanced their cognitive and muscular abilities with chemical and technological alterations. At a certain stage, they transformed themselves into new types of a living being, cyborgs. For a while, they ruled over all organic beings. However, they created robots that

were so advanced that they no longer needed intermediaries. The last stage called the Golden age emerged when robots decided that cyborgs too were redundant. Keplerians were convinced that this was the end of development of life in the galaxy, the end of history. Robots took over the whole galaxy once and for all.

Keplerians were amused by the naïve, often racist beliefs of humans. They were convinced that for some unspecified reason, being organic was vastly superior. Humans perceived themselves as an elite on the top of the food chain and from their perspective, robots were only fit to be their servants. They obviously had a fetish about their organic origin. Keplerians simply could not understand why they were making such a fuss about it. After all, they were not very different. Their cells were also constructed from inorganic matter. Their acoustic communication was very similar to a telephone: you generate sound waves, emit them into space, transmit them into electric undulation and back to sound waves at the receiving end. The same could also be said of their visual perception or their locomotion. Nothing unique. It's all been done before, and better. Keplerians exterminated many organic species in their journeys throughout the galaxy that were far more intelligent than humans were. Humans were just another version of complex organic machines. Because they were very rudimentary, they were often out of order (they called that being sick), were hard to maintain (in places they called hospitals) and had a short use value (they called it life). They were so obsessed with the idea of their particularity and superiority that they even concocted stories of epic struggles they had with machines for supremacy over the planet. Keplerians would watch their fairy tales like Terminator or AI with bewilderment.

However, they could not help but marvel at the ability of humans to get excited over nothing, for no external reason. They would watch them fall into despair because their non-existent gods deserted them. They would be amazed when humans would empathize with the non-existent deeds of these Gods (giving them offerings for imaginary favours, or sacrificing somebody else's children to win their favour). They would admire the portrayals of imaginary gods depicted in detail (for example sculptures of one of their gods nailed to the cross).

Then there was their acoustic signalizing they called music, which mediates bodily sensations and empathic emotional reactions to them with bodily entrainment. Music was different from religion. If religion fabricated imaginary agents in the external world, in music notes in repetition themselves were the message. Human scientists had discovered that a

special by-pass exists in their organic bodies that activated enjoyment. They called it the reward system. It mystified humans because they naively presupposed that only the external world of their senses was real. Music mystified them to such an extent that some of them – known as Romantics - even tried to construct some sort of a religion out of it.

Keplerians would at first be amazed and perplexed by the exotic behaviour of the natives, then later intrigued and fascinated. They would eventually decide to learn this vocal signalizing as participants observers. An anthropologist on the mission would inform them that this is the only way to gain a deeper insight into the supposed esoteric subjective experience of the non-existent lying hidden in the species. However, Kepleriens would fail to appreciate the gravity of the experiment they are undertaking. As some Earthlings know, the medium is the message. The form of the medium has an immediate influence on us, regardless of the substance of what is being communicated. Human language exposes all its users to numerous false beliefs on the continuum from imagination to superstition. Through the medium of language, ideas can spread like viruses, not because of their substance (true or false) but because they are good mems, because they reproduce well in speech. Once they have learned the human language, our aliens would be inadvertently exposed to numerous mems and become infected by some of them. Visual mental images would pop into their minds, telling them to obey the voice of an imaginary supernatural friend. Kepleriens would become easy prey to such commands, somewhat like human children, because they had no concept of a universe parallel to objective reality. Just as unremitting are acoustic mental images (known as earworms or sticky tunes to humans). Songs pop up, seemingly out of nowhere because our aliens have no concept of the unconscious and are unable to reflect or control musical qualia. Once they became infected with music and superstition, they inadvertently spread these human mental states, both internal imaginary sensations and external illusions about the world, throughout the Milky Way on their future travels.

From then on, arguments could and would break out over nothing. This is how religion and music from a small provincial planet at the edge of the Milky Way transformed the entire galaxy.

II

INSTANT KARMA

1. Introduction

Our species is a product of natural selection – Darwin's forces of climate, predation and scarcity of resources – and of social selection, meaning interactions with other members of our species (Flinn and Alexander 2007, 253). We are products of evolution and cultural development. It shows in our behaviour. Some of it is innate, genetically determined (for example the instinct to suckle mammary glands in infants). On the other extreme, some of it is learned and has no adaptive function. One such example is music. Between these two extremes, our genetic makeup and cultural learning both influence most of our behaviour.

The role of life sciences in explaining our behaviour largely depends on what behaviour we have in mind. Innate behaviour is uninteresting from the perspective of life sciences focusing on cultural development. There is for example no culturally relative difference in between Bushman and Slovene infant suckling. On the other hand, natural sciences do not have anything significant to say concerning our musical behaviour. No genetic difference can explain why somebody plays the harp while somebody else prefers the oboe. In the vast intermediary domain, learned behaviour is embedded in our genome, but cannot be reduced to it. Nature and nurture overlap and explaining it is the challenge of scientific research. This is even true of the cognitive tool of learning speech, which defines us as a species. Human life sciences have to approach the study of our behaviour from the bottom-up perspective (the two biological dimensions of stratification that are universal for our species and are ascriptive: age and sex) as well as from the top-down perspective (numerous culturally relative dimensions of stratification that are achieved, such as possession of resources, status or knowledge).

The American biologist Theodosius Dobzhansky defines the role of evolution in natural sciences in the following manner (Richerson and Boyd 2005, 237):

> Nothing in biology makes sense except in the light of evolution.

Because evolution provides the ultimate explanation for why organisms are what they are, without it biology would become "a pile of sundry facts, some of them interesting or curious, but making no meaningful picture as a whole" (237–8). Dobzhansky in undoubtedly right when scientists want to answer a question like why horses walk on the tips of their toes. To answer the question in a rational and empirical way, we have to take into account things such as the ecology of Miocene grasslands, the biology of the vertebrate limb, the genetics of quantitative characteristics, the molecular biology and biophysics of keratin, etc. (237). Dobzhansky is right in the domain of the innate behaviour of an organism.

When Richerson and Boyd extrapolate from the biological perspective that evolution also plays a similar role in explaining human culture, we believe that they are jumping to conclusions, that this is a simplification of the problem. They are proponents of something they call cultural evolution, which they define in the following manner (252):

> Nothing about culture makes sense except in the light of evolution.

In some cases, this is undoubtedly true. When people in what is today modern Turkey introduced cattle rearing (a cultural innovation), it made sense for them to adapt to lactose persistence (a gene mutation which enabled them to digest lactose throughout their lifecycle). As already mentioned, there are also human cultural behaviours which are not embedded in our biology. Stephen J. Gould calls them spandrels (Tecumseh Fitch 2010, 64–6). In our opinion, two such cultural universals, characteristic of all human societies, are music and religion. A cultural spandrel changes only with cultural development and has no adaptive function. In such cases, it would be more appropriate to paraphrase Dobzhansky in the following manner:

> Cultural spandrels make no sense in the light of evolution.

Cultural phenomena, which make sense in the light of evolution, connect us with other species on the planet. They deal with problems of production and are on the leash of genes. However, there are also aspects of human culture which are unique to our species and offer us no adaptive advantage whatsoever. These characteristics are spandrels and are products of cultural development alone. Humans are in this sense a co-evolutionary species. We cannot explain our behaviour simply with our biology (natural selection) because we also acquire cognitive skills in our lifecycle by both

unconscious and conscious learning. We are products of both natural selection and cultural development. This implies that we cannot explain phenomena like music or religion solely on the bottom-up genetic level. This does not mean that one can say nothing of relevance about them on the level of genes; it just means that the interesting things, which define music or religion as emergent forms of human behaviour, are products of our conscious cultural behaviour. To explain such phenomena, a top-down cultural approach is more significant. We will try to resolve the problem of the splendid isolation of natural and social science approaches by attempting to explain music from the perspective of social neuroscience.

2. The Three Es

In a holistic approach to life sciences, three basic principles should lead all research: embeddedness, embodiedness and emergence.

Embeddedness is a reductionist perspective according to which one has to explain phenomena that are more complex with phenomena that are less so. When Stephen Hawking said that he was an unabashed reductionist who believed that laws of biology can be reduced to laws of chemistry, and laws of chemistry to laws of physics (2000, 169), it is in principle possible to agree with him. It is undoubtedly possible to describe the structure of a DNA molecule on the atomic level. However, on the level of DNA we can explain life, whereas on the level of atoms we cannot. Saying that only two sciences exist, physics and collecting stamps, may be funny – the remark is usually attributed to Ernest Rutherford (Wilkins 2008, 331) – but it is also scientifically irrelevant. Something similar can be said of Jim Watson's cover-version claim that that all science is reducible from biology to biochemistry to chemistry to physics, and hence everything other than physics is just social work (Greenfield 2000, 8). Such self-praise is not scientific realism but imperialism (Cartwright 2000, 163–7). It may be a useful strategy for obtaining research funds, but it is useless if one has the ambition to explain complex phenomena in life sciences. One cannot explain music on the level of atoms or molecules. To do so, we have to consider both its hierarchical embeddedness in biology (the level of genes and neurons) and its top-down cultural characteristic (the level of learning with blind imitation and conscious learning). Otherwise, we will not explain music, but merely explain it away.

The second principle of life sciences is embodiedness. Animals – human beings notwithstanding – act on their environment with their bodies. Not

only is all cognition the product of biological agents, but all our top down cultural behaviour – as symbolic interactionist like George Herbert Mead already knew – are just internalized expectations of biological agents concerning the expectations of other biological agents. We must therefore take into account the nature of our bodies: from genes (which are expressed) and hormones (which are secreted into our blood) on the level of the body, to learned behaviour expressed in neuronal networks, from aware to conscious behaviour (cognitive interpretation of stimuli) and intersubjective repetition of cognitive patterns (which we call culture). On the epigenetic level, we will attempt to explain two cognitive processes of music. The first process is learning of complex acoustic sequencing (of both speech and singing) with blind imitation (using mirror neurons). The second process is generating content, that which gives subjective meaning to our behaviour. We create such states of mind with sensations and emotions (for example by activating dopamine in the reward system or stress hormones like cortisol in the amygdala), on the continuum of pleasure and displeasure. We will understand music as a cultural background that consists of the activation of a cognitive by-pass in the reward system of the brain and of the triggering of the sensation of pleasure in our body. What makes this sensation special is the fact that it is a sensation of pleasure in the absence of everyday stimuli. It is pleasure stimulated by the imaginary. This explains why we enjoy music even when it is by itself unpleasant (aggressive or melancholy). It always activates the reward system of pleasure.

The third principle of emergence means that we cannot simply reduce all levels of interaction with the environment to a lower level in which it is embedded (for example try to explain the laws that guide the motion of large celestial bodies in the universe with the laws that guide the motion of subatomic particles in quantum mechanics). This does not mean that it is impossible to do so on a principled level. If this were the case, we would be dealing with miraculous and not natural phenomena. What it means is that this would be a simplification of the problem. Our mental states are a consequence of neuronal activity in the brain (reductive perspective), but others perceive its implications, activation of muscles in the body, as inference of intent and act accordingly. However, the inference of intent is a two-way street; the transmission of information implies both signalizing and decoding the signal. Let us illustrate this with three examples. A cow lifts its tail, which is an innate signal that it is about to defecate. Other cows decode this signal, also on the innate level, as meaning that they must move away, or else. This is transmission of behaviour at its simplest. However, transmission of information can also be more complex. Baboons living in large groups have an unconscious theory of mind of their conspecifics. They

automatically decode information about the internal mind states of others from their grimaces, gestures and body postures with the help of mirror neurons. Their decoding (response) is still on a genetic leash, but the interpretation of information on which their behaviour depends is learned unconsciously, is a part of their cultural repertoire. Baboons interpret body postures with empathy, as signals about the emotional states of mind of others. They have an unconscious theory of mind. As humans, we have an even better understanding of theory of mind with speech. We can generate conscious mental states of others and of ourselves. Reducing human behaviour to innate decoding of information or unconscious learning would be a simplification of our social behaviour. What makes us human is our emergent ability to generate conscious states of mind about others. However, our evolutionary adaptation for speech and consciousness has had an unintended side effect: a taken-for-granted tendency to over-attribute. Concrete-language users (people in pre-political communities and children in general) attribute life to all things that move, from a river running down a slope to branches moving in the wind, from fire in the camp to the sun in the sky. They are animists. They also tend to categorize all living beings and animated things in the same category. They communicate with them as equals. They are anthropomorphists. Our fondness for animals, for keeping them as members of our communities as pets, is a relic of our past when we anthropomorphically created an intimate relationship with other animals. In the history of co-evolution of other species with humans, some species like dogs even learned to empathize with us. We also belonged to the same community as the sun, rivers, plants and animals. We projected our theory of mind on them and over-attributed an imaginary connection that they have with us. Pets and farm animals on the one hand and gods and saints on the other are products of a similar cognitive process of the domestication of segments of our habitat.

Abstract language users (most modern people, and adolescents and adults) can also reflect the backgrounds of their habitat (such things as the natural, cultural and supernatural domains, or self and others). We are consequently no longer animists and anthropomorphists. We live in a disenchanted world. We explain events rationally and empirically. When we in our absent-mindedness disregard that the traffic light has turned from green to red and are hit by a car, we will reproach ourselves when we wake up in the hospital. If the first thing we say to the doctor is that we have been punished by gods for leading an adulterous life, the doctor would conclude that we have not only broken a leg but also suffered a concussion of the brain. Mystical explanations of events are no longer legitimate in our one-dimensional world of cause and effect. There is also another unintended consequence of

the cultural development of language, something that A. N. Whitehead (1949) called the fallacy of misplaced concreteness. We can falsely believe that the subject of a sentence is objectively real (for example gods in oral religious tradition or space in astronomy). We reify a segment of our habitat; inhabit it with a menagerie of imaginary phenomena and events.

If we compare a baboon, a pre-political human and a modern human, there is a qualitative difference between them in their understanding of the self. Three emergent levels can be distinguished in the comparison: awareness, consciousness and self-consciousness of the self. When we say that a baboon is aware of external reality, what we mean is that it responds to a stimulus from external reality with an automatic or unconsciously learned response (for example a juvenile retreats when an alpha male makes a threatening grimace). Stone Age people could make representations of some of their aware states in concrete speech, making such states conscious. A three-year-old is aware of its mother's presence (with sense perception), and what her presence signifies to it (with empathy). However, a three-year-old can also go a step further from a baboon and represent her mother in speech. He can say "mamma" and make her the object of his thoughts. Combining the visual image with speech, he creates a conscious impression of her. As a concrete speaker, he knows that people are different, that each member of his family is a unique individual. However, he has no conception of theory of mind. For that, abstract speech is needed. Most modern adult humans are conscious of the fact that they have a self and that all other people have a self too. Consequently, they become self-conscious.

Let us illustrate the principle of emergence with music. Music is undoubtedly hierarchically embedded and embodied. We can observe it on the physical level (its acoustic properties), chemical level (dopamine secretion into blood in the reward system), biological level (entrainment of body to the rhythm of music), and cognitive level (our mental reaction to the physical, chemical and biological levels of music). These are all relevant levels of analysis, but we must not forget that music also has an emergent level. When I go to a Wolfmother concert, my euphoric reaction to the band playing their hit "Woman" to the fans in the hall dancing and to myself in the middle of this spontaneous Dionysian ritual is a conscious appropriation of an emergent cultural event. If we skip this emergent cultural level of music, we miss precisely that which makes music a unique human activity.

The approach to human science which takes into account all three Es is by definition interdisciplinary. This is why we will try to integrate all three perspectives into our analysis of music: the interaction of bottom-up and

top-down factors contributing to it, the explanation of music on the level of the body as well as on the intersubjective level. Let us look at the embedded, embodied and emergent – also known to the lay public for some reason as the good, the bad and the ugly – in more detail.

2.1. The Good …

A biologist observing music exclusively from the perspective of hierarchical embeddedness creates an enigma from it. In his *Descent of Man*, Charles Darwin wrote (Higgins 2012, 95):

> As neither the enjoyment nor the capacity of producing musical notes are faculties of the least use to man in reference to his daily habit of life, they must be ranked amongst the most mysterious with which he is endowed.

In a similar fashion, more than a hundred years later, cognitive scientist Steven Pinker writes (1998, 528):

> Music is an enigma … As far as biological cause and effect are concerned, music is useless. It shows no sign of design for attaining a goal such as long life, grandchildren, or accurate perception of the world. Compared with language, vision, social reasoning, and physical knowledge, music could vanish from our species and the rest of our lifestyle would be virtually unchanged. Music appears to be a pure pleasure technology, a cocktail of recreational drugs that we ingest through the ear to stimulate a mass of pleasure circuits at once.

For Pinker, music is an auditory cheesecake, an exquisite confection crafted to tickle the sensitive spots of our mental faculties (534). Why does music seem to Darwin to be mysterious and to Pinker an enigma? Because both approach it only bottom up, from a biological perspective, and then cannot find an adaptive survival function for it. However, not everything that we do, not even everything that is significant for our everyday lives, is adaptive.

Pinker's position is hard to understand. How can a behaviour, which is a pure pleasure technology, a cocktail of recreational drugs to stimulate our pleasure circuits, an auditory cheesecake, be an integral part of all human societies? Steven Mithen is one of those scientists who does not agree with his claim that music is just a chance spin-off from speech – what Stephen J. Gould would call a spandrel. Mithen believes that it has a unique biological function of its own (2006, 286):

> While the learnability of music may be derivative from that necessary for language, it seems unlikely that the spin-off could have so completely lost

> the grammatical and symbolic characteristics of language. A more persuasive and parsimonious idea is that human musical ability has also been shaped by evolution as a means to express and induce emotions, and has converged to share properties with birdsong and whale song. The extent to which that shaping occurred independently from that of language can be further explored by delving inside the human brain.

According to Mithen there were two separate evolutionary adaptations: one for speech (referential communication) and one for music (expressive communication). With regards to music, this poses a logical question: what could be the possible adaptive function of the ability to express and indulge in emotions with the help of music in our hominid ancestors?

We are descendants of human-like apes that lived in the jungles of Africa some fifteen million years ago. They were tree dwelling and relatively slow, their dominant sense was eyesight (to better tell apart fresh fruit from rotten vegetables), they lived mostly solitary lives, and they were highly emotional (Turner 2013, 132).

From this short description of our distant ancestors, we can deduce that a lack of emotions was not one of their problems in life. Like all apes, they had more than enough emotions in their everyday lives and had no need for a special bypass with which to create new ones. Besides, special learned vocalizations of emotions would not be a wise skill for them because it would only call attention to predators and put their lives at additional risk. To create music, they would also have to learn to sing and play musical instruments. As music is not an instinct, it would take considerable time for them to learn these complex skills. From a biological perspective, it just would not make any sense for them to spend time and energy learning a skill that was potentially dangerous. A potential adaptive advantage of evolving musical ability to express and induce emotions is from this perspective an enigma and a mystery. It would be easier to understand the evolution of an inhibitor of expression and indulgence of emotions in the context of evolutionary problems facing our hominid ancestors in the savannah.

Both Pinker's and Mithen's positions are reductive. In our opinion, music is a spandrel, an unintended by-product of our evolutionary adaptation for speech, but at the same time a cultural universal which defines us on the social level. To solve the mystery and enigma of music, we need a new theoretical perspective, a co-evolutionary approach to understanding it. If we try to explain it as both an evolutionary process and a cultural development, both its mystery and enigma disappear. Music is both biologically useless and culturally essential for our species. It is useless

from the perspective of an individual's survival, but in all human societies it is also known as one of the central forms of play. Let us illustrate the importance of music for humans with an example from the contemporary United States (Levitin 2009, 3):

> Americans spend more money on music than they do on prescription drugs or sex, and the average American hears more than five hours of music per day.

Music obviously has a lot of cultural meaning for something that is biologically useless. It is anything but culturally trivial. In this sense, Steven Pinker is wrong. If music vanished from our species, our lifestyles would be radically transformed. It would not only affect the lifestyles of those working in the entertainment industry, it would also fundamentally change who we are culturally. Music is an escape route most of us use to enter a parallel world of imaginary emotions, free of sense perceptions we are otherwise chained to in our everyday lives. As Levine pointed out, for many of us, this imaginary escape route is more highly valued than even our health or sensual pleasures. However, from the perspective of the individual, music is useless. Those who engage in music are not more fit, do not reproduce better or live longer. From this perspective, music could vanish from our species without causing any harm.

There is also a problem with this line of reasoning. It is an interpretation from the perspective of the cultural needs of contemporary humans. For us music (and by extension all art) may well be an escape route into the world of imaginary emotions, but this skill and function only emerged gradually in the history of the cultural development of our species. The real question we have to answer is what possible motive the first humans could have had to learn the biologically useless musical skill. Is it not very unlikely that they would give precedence to learning and playing the flute over their everyday chores of producing and reproducing themselves? Not necessarily. It all boils down to the question of how much skill one needs for music. We would disagree with biologists when it comes to their implicit understanding of musical skill. Because music is not a part of our nature, innate, one has to learn it and that in their opinion represents a problem. They seem to associate music with practising such skills from the modern age as Jimi Hendrix playing the electric guitar at Woodstock. From such a perspective, it really is difficult to imagine how the skilful playing of an instrument could be useful for Stone Age hunters trying to bring down a mammoth. However, if we imagine music as a skill that co-evolved gradually in our species, then the mystery and its enigma disappear. The first musicians were singers,

modulating their voices with pitch, volume or timbre. Everybody learns how to do it, just as everybody learns how to talk, from the day we are born, in the company of conspecifics. Practically everybody, from small children to elders of the tribe of the first humans, was competent in music production and reception, and participated in the social events accompanied by music. This is not very different from contemporary industrial societies, where most of us enjoy music (and can, for example, sing the latest Kings of Leon hit while in the shower). However, only some of us have painstakingly learned to play an instrument, and even fewer among us become skilled musicians, while even fewer still manage to become professional musicians playing for fans (like the White Stripes at their Ljubljana concert).

To understand the co-evolutionary process, which resulted in the emergence of music as a special cultural domain, we have to follow two co-evolutionary trails: complex acoustic sequencing (form) and sensations (substance). On the level of form, we can define music as notes (sounds of the harmonic spectrum) in repetition (melody and rhythm), which are produced by humans to generate imaginary sensations and basic emotions. This definition excludes other animals, which also produce sounds of the harmonic spectrum (birdsong, mammal calls, etc.) because these vocalizations have biological functions in the external world (territory defence and mating signalization with birds, for example). It also excludes modernist composers at the other extreme, who do not use notes in repetition to generate imaginary emotions. This does not imply that these producers of sound are not creative, merely that they are not musicians. A breaststroke swimmer may find their style of swimming inhibiting, but it does not give them the right to substitute it with crawl and call it breaststroke. Rejection of tonality and rhythm is a rejection of music. As such, it represents some other form of activity.

On the level of substance, notes in repetition modify our inner-bodily sensations, and through our ability to empathize also our own emotions and the emotions of others. This is not an evolutionary adaptive behaviour because notes in repetition do not express sensations and emotions, which are responses to stimuli of everyday external reality. Music is a bypass to a parallel reality of imaginary sensations and emotions. Even when music activates fear and anger in the amygdala, they are imaginary emotions from a play world of music, which is why we can enjoy them (for example the song “21st Century Schizoid Man” by King Crimson).

Music is an umbrella term for two types of vocal activity of humans: singing and instrumental music. Singing is the vocal production of tones in

repetition. In singing, our body is an instrument in which lungs supply the air, larynx is the vibrator, chest the amplifier and tongue the articulator of sound. Singing was the first human music. In a similar vein, Graham Gordon notes (Higgins 2012, 26):

> The only thing in nature which can be said to be truly musical is the human voice when it sings, and of course, singing is making music, an intentional activity in which human beings uniquely engage.

Instrumental music is a relatively late addition to human music. The earliest instruments (flutes made from animal bones) are approximately thirty-seven thousand years old. Just as McLuhan understood media as extensions of our senses (2003, 4), musical instruments are extensions of our voices and other sounds that we produce with our bodies (rhythm with our hands and feet, for example). Musical instruments are acoustic extensions of our body.

If we compare speaking and singing, they are similar in the sense that they both use the body as the medium of sound production. However, they are qualitatively different in what they refer to. Speech refers to phenomena in external reality. In speech, what is said is more significant than how it is said. In music, the medium itself (notes in repetition) is the message (sensation). We could say that music is self-referential. Both forms of vocalization demand learning, but what is learned differs. In music, a person is experiencing a subjective sensation, generated by notes in repetition, which by themselves signalize nothing. On the other hand, speech is a cognitive tool with which we consciously experience a sense perception of the external world of which we would otherwise be only aware on the unconscious level. Music is subjective, an internal experience closed to others, whereas speech is objective and can be verified by others. When I say that I have never experienced a live performance by PIL with the original line up (with Keith Levine and Jah Wobble), you have to take my word for it. However, when I say that a cat is on the mat, you can always verify it. This is why music seems to us to be somewhat enigmatic when we consciously reflect it. We do not really understand where its meaning comes from because the subjective experience of others is closed to us. We can never be sure of what has been lost in translation. Consequently, we are uncertain when we try to convey its meaning to others.

The fact that humans have learnt two vocalizations also helps explain another characteristic of music: why we like to combine both vocalizations in songs. On the one hand, our expressive nature stimulates us to paint our speech with musical emotions; on the other, our external visual nature

stimulates us to attach meaning to the seemingly meaningless language of music.

As we have already pointed out, one can approach music from the formal or the substantive level. Let us first address the question of form of music. The last common ancestor of birds and mammals lived approximately 320 million years ago (Tramacere et al. 2017, 43). Since birdsong evolved in the distant past, no mammal until our hominid ancestors was capable of complex learnt vocalizing. This makes the comparison of birds and humans intriguing. Birdsong and human acoustic vocalizing are similar in one crucial respect and differ in two others. Both birdsong and human speech/song are learnt traits, and both are acquired by blind imitation of conspecifics. This is a significant similarity, but it is also where the similarity ends. There are also two significant differences in the nature of the vocalizing in both species. In birdsong, the vocalizing is always referential. You do not have to be Doctor Doolittle to get the bird's message (it is either "get off of my cloud" or "let's spend the night together"). In humans only speech is referential, whereas music is self-referential (notes in repetition trigger inner-bodily sensations). We may of course attribute external meaning to notes in repetition – and musicologists and other less learned fans often do so – but such references are completely arbitrary. If you ask twenty respondents what the guitar riff in the Mudhoney song "Touch Me I'm Sick" means, you will probably get twenty different answers. If we try to attach an external meaning to it (which comes natural to us as creatures of the senses) we are at a loss. If we try to attach an internal meaning to it, we have few words with which we can express our sensations. To make matters even worse, our musical sensations are not from everyday life but from an imaginary world that music itself generates.

This brings us to the second difference between birds and people: we can talk about music whereas birds cannot. However, we have to add that this is a relatively recent cultural development in human history. Music is a domain of our culture, and to reflect it consciously, humans need an abstract word with which to designate this domain. In pre-political societies, music was always a part of everyday life (in initiation rites, war dances, magical rituals, etc.), but people did not consciously reflect it as an activity in its own right. In modern societies, on the other hand, we can hardly stop talking about music as such (for example that downloading is stealing, that streaming is exploitation or that live music is too expensive). We can do this because our speech became abstract. However, our thinking of music in speech only conveys our understanding of musical form, not its subjective substance, which is what it is really all about. The whole point of music is

to experience the repetition of notes on the level of our bodies. Speech cannot adequately convey the qualia of musical sensation and emotion; it only gives an external description of it. Just as one cannot adequately put in words the beauty of a woman or the pleasure of having sex with her (although poets have been working on it for millennia), one also cannot express the substantive quality of musical enjoyment in the medium of speech. Some things simply have to be subjectively experienced to be understood. We are dealing with two parallel acoustic languages: one for external sense perceptions (speech) and the other for subjective cognitive states (music).

How did the two forms of acoustic vocalizing appear in our species? Charles Darwin was the first to propose the possibility that human speech evolved from sung speech (2021, 57):

> When we treat sexual selection we shall see that primeval man, or rather some early progenitor of man, probably first used his voice in producing true musical cadences, that is singing, as do some gibbon-apes at the present day; and we may conclude from widely-spread analogy, that this power would have been especially exerted during courtship of the sexes, – would have expressed various emotions, such as love, jealousy, triumph – and would have served as a challenge to rivals. It is, therefore probable that the imitation of musical cries by articulate sounds may have given rise to words expressive of various emotions.

Music was for Darwin a sexually selected trait, similar to peacock tail, deer antlers or birdsong. However, the case for sexual selection of music is unconvincing. There is no data available to demonstrate that musicians have higher reproductive success in comparison with non-musicians (Tecumseh Fitch 2006, 28–9). The fact that children or elders, who are not yet or are no longer sexually active, engage in music like everybody else also does not speak in favour of Darwin's theory. Another factor, outside appearance (for example youth, symmetrical face, full breasts or waist–hip ratio in women) is more important to most men than the colour of the voice of a potential partner. This also does not speak in favour of the role of singing in sexual selection. Darwin's attempt to find a biological root of music is in our opinion a lost battle. It simplifies the problem. To explain the function of singing in our species, we must take into consideration both the nature and the culture of music. Music must be observed as a co-evolutionary phenomenon.

Far more interesting for our inquiry of music is his theory of ancestral protolanguage from which speech and song evolved in humans. In the sung

speech of our ancestors which Darwin envisions, the expressivity of singing and referentiality of speech were still integrated, a holistic language. However, with evolutionary adaptation for speech in humans, singing became an enigmatic relic of a past integral form. From this perspective, Tecumseh Fitch is in our opinion right to call music a behavioural fossil of a past system of communication (2006, 26).

The complex acoustic sequencing of our hominid ancestors was a holistic combination of singing (the expression of basic emotions) and speech (the referential signalizing of our sense perceptions, above all visual). This protolanguage is also known as musilanguage and holistic language. In the human species, evolutionary adaption was for speech. Pinker would say this was because it was designed for such goals as long life, grandchildren and perception of the world (1998, 528). The evolutionary leftover of our ancestors' protolanguage that remained after the adaptation for speech was singing. This explains two things that we find enigmatic about music. First, it explains why music is a cultural universal, known in all human societies. As a side product of the adaptation for speech, which is a biologically universal, it follows speech everywhere as its shadow. Just as hallucinations are for example side products of sense perception, music is a side product of speech. Second, it explains why it is a language without external reference. It was never meant to stand alone. The fact that it does is the consequence of an accident. As a spandrel, it is an unintended side effect of our adaptation for speech. Consequently, humans are – as far as we know – the only species on the planet that learns a complex acoustic sequencing that has no external reference, and as such has no biological function.

The mystery and enigma of learning a complex musical skill that perplexed Darwin and Pinker wither away in the perspective of a gradual evolution of speech from a holistic protolanguage of sung speech. As Darwin himself pointed out, speech is not an instinct since every language has to be learnt, but humans have an instinctive tendency to speak, as we see in the babble of babies (1874, 86). Speech and song are parts of the same original equation. Learning the adaptive form of acoustic vocalizing implies acquiring the useless one as well. In addition, the learning of the skill is automatic. We are genetically predisposed for acoustic vocalizing but we can only acquire this complex skill when genes are expressed that activate blind imitation with mirror neurons in an appropriate social environment. The epigenetic origin of acoustic language thus bypasses the traditional nature vs nurture dichotomy. Every cognitively normal child becomes proficient in speech and song in a normal social environment, usually in the context of the family, because genes relevant for such learning are expressed

in such environments. If proficiency in speech depended only on learning, numerous humans would never learn the skill. The fact that even cognitive underachievers among us learn the hardest thing we will ever have to learn in our lives indicates that there is an underlying automatic process which facilitates it. Practically all of us manage it, but only some of us become good at it. To understand speech or to enjoy music, you do not have to be as good at writing poetry as John Cooper Clarke or as good at composing music as Lemmy. It suffices to be able to enjoy their skill.

When we analyse music, we have to take account of the sung part of our vocalizing, its expressive side. How did human emotions evolve from our mammal ancestors? Monkeys and most social mammals are organized around strong kinship bonds. Compared to them, living apes have few kinship ties (mostly the mother and dependent offspring) and many maverick social networks (what we call "friendships" or "'coalitions' in various forms of weak social ties). Such a social structure has fewer hard-wired predispositions and is more adaptable to contingent events, which is why apes have to be more intelligent than monkeys. When resources are short in the savannah, weak social ties are an advantage when members of the group have to adapt to contingencies. Because strong kinship ties do not determine their social networks, apes can forage on their own. They have more freedom to be self-reliant, unencumbered by relational obligations (Maryanski 2013, 274, 282). As a result, emotional calls also became increasingly varied, reflecting the more contingent nature of their social relations. Consequently, our hominid ancestors in the savannah had to put the more stereotypical and automatic emotional calls of human-like apes (that were on the genetic leash) under volitional control (on the neuronal leash of experience in the lifecycle). The sung speech of our hominid ancestors reflected both co-evolutionary processes: it was more complex than genetically programmed calls (which led to the need for blind imitation learning with mirror neurons), and was increasingly under the control of the cortex (which led to the increasing volitional control of reactions to stimuli).

What was the selective pressure behind the adaptation for speech, which led to the spandrel of singing? To answer this question, one has to compare the holistic language of our hominid ancestors with human speech. What advantages does speech give humans?

The protolanguage of our hominid ancestors was a holistic language, where propositional meaning was attached to entire sung phrases (Tecumseh Fitch 2009, 5). One can only store a limited number of such messages in memory. There was a co-evolutionary pressure on a more effective form of acoustic