

# How Music and Our Faculty for Music Are Made for Each Other

JAMES PARAKILAS

**Abstract** This study relies on the prevalence of certain structures that largely distinguish the creation and reception of music from that of language – namely, temporal grids, scalar grids, and segments with their repetitions – to construct a model of the human cognitive faculty for music that allows humans to make music the way they do. The study draws on research and thought in philosophy (including phenomenology), linguistics, psychology, and neurology, coupled with musicology, to produce a model of a human capacity to make complex comparisons between ongoing sound sequences and those simultaneously reconstructed from memory by registering the relativities within their flow. This model is then used in a consideration of how the faculty for music interacts with the faculty for language in the experience of song and a consideration of how a similar cognitive capacity for music might be identified in other species.

Nick Pyenson, a leading whale researcher who is curator of fossil marine mammals at the Smithsonian Institution, begins a popular-science book about whale archeology and biology with something calculated to make the subject tantalizing to his readers: he evokes the ‘complex songs’ of male humpbacks, ‘composed of phrases collected under broader themes, nested like Russian dolls that repeat in a loop’. Then, two pages later, Pyenson writes that these whales ‘speak to one another with impenetrable languages’.<sup>1</sup> That might not seem like a big change in the terms of description. After all, songs normally contain language, and there are deep overlaps between music and language, even when they are isolated from each other, that facilitate their combination in song. But describing humpback whales as our fellow singers may invite one kind of wonder and describing them as speaking, even speaking ‘impenetrable languages’, may invite a very different kind. After all, we do also habitually make a distinction between *music* and *language* (or *words*), just as we make a distinction between *song* and *speech*, both in academic discourse and in everyday situations: ‘words by Gilbert, music by Sullivan’ or ‘words and music by Stevie Wonder’. Speakers of English (and other languages that use a word derived from the ancient Greek *mousikē*) seem to have little trouble thinking of music and

---

Email: [jparakil@bates.edu](mailto:jparakil@bates.edu)

I thank Jeremy Day-O’Connell, Gina Fatone, Mary Hunter, Daniel Leech-Wilkinson, Ralph Locke, and the reviewers for this journal, all of whom have given me valuable responses and suggestions.

<sup>1</sup> Nick Pyenson, *Spying on Whales: The Past, Present, and Future of Earth’s Most Awesome Creatures* (Viking Penguin, 2018), 1, 3.

© The Author(s), 2024. Published by Cambridge University Press on behalf of the Royal Musical Association

language in one moment as nearly interchangeable and in the next as terms that define a difference.<sup>2</sup>

Pyenson's shifting terminology notwithstanding, the researchers who have recorded and studied the humpbacks' elaborate sound productions for over half a century have quite uniformly labelled it *song*. We might imagine that to be a sentimental choice, but in fact it rests on one of the earliest achievements by those researchers, which was to discover features and patterns in the whales' strings of sound that lend themselves to analysis in the ways that music scholars have long analysed the formal patterns of melodies. But if we want to know what kind of phenomenon we are dealing with in those sound productions, we would presumably want to know more about them than what kinds of patterns they form. We might well ask how those 'musical' sound patterns connect to the singers' life situations when they are singing, what it means that they share their songs with other members of their group, and how their minds and bodies are engaged in producing and reproducing their songs. Unfortunately, observing any such things about these giants of the deep is extraordinarily difficult. It has taken immense ingenuity over the course of decades even to discover the mechanism by which they produce their sounds. And there is another difficulty.

When we ask what kind of phenomenon humpback 'song' is, we are implicitly seeking comparison with phenomena we understand better, which would naturally be human forms of sound production such as music and language. In those cases, researchers have endlessly studied the formal patterns, the social and emotional functions, and the cognitive processes of production. Furthermore, in all those kinds of study, they have distinguished music from language. Linguists, for instance, have proposed that humans have a set of skills that constitute a cognitive faculty specifically adapted for producing and apprehending language, while theorists of music cognition posit a cognitive faculty (which some call *musicality*) specifically adapted for producing and apprehending music. In neither case has a single, definitive, and uncontested model of that faculty emerged. But there is agreement that these are two different faculties, and that is significant because a distinct model of, say, the human faculty for music could point us towards the qualities of human music that make it a significantly different phenomenon, or experience, from language, for all their overlapping qualities. But that pointing is exactly what we are lacking because we have not had an account of how our faculty for music is geared to producing a phenomenon that has the qualities of human music – and conversely how the qualities of human music reveal the nature of

---

<sup>2</sup> Languages of many other cultures use terms both broader and narrower than the European word *music*, rather than any term equivalent in its coverage, even if ethnomusicologists say that every culture has something that speakers of English might want to call *music*. See Bruno Nettl, 'Music', *Grove Music Online*, 2001 <[www.oxfordmusiconline.com](http://www.oxfordmusiconline.com)> (updated and revised 1 July 2014), II and II.7. But then, the European tradition itself was ambiguous from the start: Plato used the word *mousikē* to refer to the whole realm of the Muses, which included language and dance and other activities beyond what we call *music*, and yet he also wrote of it as the art (*tekhnē*) of modes and rhythms, the elements most defining of *music* in our narrow sense. Compare Plato's language about the Muses in *Phaedrus* 259b–d to his language about harmonies and rhythms in *Alcibiades I* 108c and *Republic* 398c–400c.

the cognitive process by which we create and experience it. The purpose of the present article is to propose such an account.

For that purpose I set out a model of the human faculty for music as a set of basic cognitive capacities apparently shared across the human species and no doubt applied by us in a variety of tasks, but particularly suited for, in fact tested to their utmost in, our production and experience of music. Then I consider what there is in human music that would facilitate, and demonstrate, the exercise of that faculty. Here – assuming we accept that there are arts or cultural behaviours that can be connected across cultures globally and historically as *music* – the challenge is that those arts are so diverse as to make it difficult to isolate characteristics as musical ‘universals’. Two solutions that have been proposed to meet this challenge are at work in the present study. One is to consider a number of ‘statistical universals’, characteristics not all of which are necessarily present in any one kind of music, but some combination of which are always present.<sup>3</sup> Another is to think of musical universals not so much as features of the musical product, but as characteristic processes of music making.<sup>4</sup> By that way of thinking, the cognitive faculty for music could even be what is most universal about human music.

To think of music in terms of its universal (or universal enough) features or processes focuses attention on the ways music is formed, but that does not mean that music can be defined simply as patterning in sound, when it is clearly, from other perspectives, a way humans relate body to mind, relate themselves to others, relate their interior processes to the world outside themselves, relate the immediate to the past and the remote, and draw real emotion from the process of fabricating with sound. The proposal being set forth here focuses on the cognitive processes that lead humans to form their music in a characteristic way because, I believe, that characteristic form of expression is precisely what triggers other cognitive processes – the ones involved in generating those relational and emotional experiences that identify what music is in our lives. I focus here solely on the relationship between one cognitive faculty and the ways we form our music; it is beyond the scope of this study to explain how exercising that faculty on music formed in that way might trigger all the things that give music meaning to us, though I point along the way to some places where we might look for those explanations. Nevertheless, because music and language are so universally intertwined in human song, I do also consider here how the music faculty and the language faculty might operate in our minds simultaneously when we make song. And

---

<sup>3</sup> See Patrick Savage and others, ‘Statistical Universals Reveal the Structures and Functions of Human Music’, *Proceedings of the National Academy of Sciences USA*, 112 (2015), 8987–92. The features of music described in my study have considerable overlap with those that Savage and colleagues propose.

<sup>4</sup> See Catherine J. Stevens and Tim Byron, ‘Universals in Music Processing: Entrainment, Acquiring Expectations, and Learning’, in *Oxford Handbook of Music Psychology*, ed. Susan Hallam and others, 2nd edn (Oxford University Press, 2016), 20–32. The processes described in my study have overlaps with those Stevens and Byron propose, especially in the realms of response to temporal structures and the processes of segmenting and grouping. The main differences lie in their concern with emotional response to generic structures, something I do not discuss, and their lack of recognition of the comparative processes that I identify as essential to the cognitive faculty for music.

I then return to the case of humpback song to ask whether studying the correspondence between the faculty and the phenomenon of music in humans can help us deduce anything about a cognitive faculty of another species from the formal patterns of its song.

## A Model of the Human Faculty for Music

The cognitive processes engaged in making and listening to music have been investigated within many fields of inquiry, including ancient and modern philosophy, cognitive psychology, neuroscience, biology, and various kinds of music study. I bring work from all these fields into contact here, conscious that each of these fields has its own train of thought, research, and debate on the subject and has developed a more or less discrete set of questions, terminologies, and methods of investigation. My intent is not to suggest that findings in one field will necessarily answer questions raised in another – not to suggest, for instance, that brain structures discovered by neuroscientists will in themselves explain psychological processes, let alone resolve philosophical disputes. Rather I draw on the work being done in various fields – and on what has been done in various periods – in order to compare the benefits of their perspectives and in hopes of providing a refreshing challenge to the boundaries of inquiry that grow up in any one field.

On that premise I begin my search for a model of human musical cognition with a source that predates the modern psychological idea of cognition, in fact the modern concept of science, by millennia. It is nevertheless, in its own refreshing terms, a thoroughly cognitive concept of musical listening that we find expressed in this passage from the *Elements of Harmony* written around 300 BCE by Aristoxenus of Tarentum, a leading philosopher of Aristotle's school and the most important ancient Greek music theorist:

It is clear that understanding melodies is a matter of following with both hearing and reason things as they come to be, in respect of all their distinctions: for it is in a process of coming to be that melody consists, as do all the other parts of music. Comprehension of music comes from two things, perception and memory: for we have to perceive what is coming to be and remember what has come to be. There is no other way of following the contents of music.<sup>5</sup>

This formulation seems so simple and self-evident that it would be easy to miss how precisely it captures what is distinctive about the cognition of music, as well as how much it leaves for us to fill in.

Let us take the precision first. We use perception in combination with memory in everything we do, not just in experiencing music. But note that Aristoxenus adds: 'There is no other way of following the contents of music.' If we do not read that as simply a rhetorical flourish, we can see that it shows how the process he is describing

---

<sup>5</sup> Aristoxenus of Tarentum, *Elementa Harmonica (Armonika Stoikheia)*, trans. Andrew Barker, in Barker, *Greek Musical Writings*, vol. 2: *Harmonic and Acoustic Theory* (Cambridge University Press, 1989), 15.

pertains with particular force to music. Take the idea of ‘following’: Aristoxenus’ verb *parakolouthēin*, like the English ‘follow’, encompasses both ‘tracking the course’ and ‘catching the drift’. ‘Following the contents of the music’ can mean tracking the course of the music, that is, tracking the succession and flow of its sounds. And it can also mean catching the drift of the music’s content. How to understand what constitutes the drift – the sense – of music is a notoriously difficult question.<sup>6</sup> But there can be no doubt that catching the drift of music is inseparable from tracking its course – that its sense resides in its particular succession and flow of sounds. What kind of sense is that?

It is not like linguistic meaning, derived from the reference each word makes to things outside the medium of language. Instead, in Mark Johnson’s account, ‘music is meaningful because it can present the flow of human experience, feeling, and thinking in concrete, embodied forms – and this is meaning in its deepest sense’.<sup>7</sup> In Daniel Leech-Wilkinson’s formulation, ‘the dynamics of the succession of sounds cause them to seem alive and construct a vivid sense of “now” that seems to move with them through time’.<sup>8</sup> All the arts create a sense of what it is like to be alive, but music is distinctive in the way it turns us, whether we are performing or listening, into the experiencing subjects whose seeming vitality we are tracking. As we track the rhythms and pacing of its sounds and silences, we feel an impulse to move, especially in our limbs, and a variety of our bodily systems get attuned to its flow, so that in ‘following’ the music, we may feel that it is coming from within us. As we track the particular course that the phrases of any music take (what Leech-Wilkinson calls the musical ‘shape’ that models the shape of a human experience), we undergo feelings whose physical manifestations (including at times tears, chills, or swooning) may correspond to what we take the character or mood of the music to be – or may seem at odds with that character, as when we cry at the sweetness of music. We find ourselves drawing successions of associations, images, memories, and fantasies from our experience, and these may be related to the feelings we are simultaneously experiencing – or may seem unrelated, as in a dream. And as our tracking absorbs the course of the music into our consciousness, we may feel an intensified connection to those around us and to our surroundings – or an intense detachment from our here and now, or even from our sense of who we are.

Do we track other arts in the same ways, or in the same complex of ways? We come closest with dance, which we generally experience in conjunction with music. But though we experience a conjunction of language and music in song, language in itself does not require us to cling to the flow, or even the order, of its events the way music does; in consequence, we are not apt to feel the liveness of language as coming from within us, as we do with music. Tracking the course is so essential to the experience of music that we can transpose, vary, or arrange it without losing track of its identity and

<sup>6</sup> See the survey by Ian Cross and Elizabeth Tolbert, ‘Music and Meaning’, in *Oxford Handbook of Music Psychology*, ed. Hallam and others, 33–45.

<sup>7</sup> Mark Johnson, *The Meaning of the Body: Aesthetics of Human Understanding* (University of Chicago Press, [2007]), 236.

<sup>8</sup> Daniel Leech-Wilkinson, ‘Musical Shape and Feeling’, in *Music and Shape*, ed. Daniel Leech-Wilkinson and Helen M. Prior (Oxford University Press, 2017), 358–82 (p. 363).

sense. But when we translate or paraphrase language – procedures that we cannot perform on music – we alter the succession and flow of the words and yet preserve their sense.<sup>9</sup> With music, then, catching the sense is inseparable from tracking the course, and when Aristoxenus writes that ‘there is no other way of following the contents of music’ than with perception and memory, we can understand him to be telling us that in using those two means in combination to track the course of music, we are at the same time using all the means we have or need in order to make sense of it too.<sup>10</sup>

Now let us consider what his model of music cognition leaves us to fill in. The most obvious question he leaves to us is, ‘How do we employ memory in relation to perception when we are experiencing music?’ No doubt we compare one to the other, but how? One possibility is that we register how each new moment of the music we are perceiving relates to the remembered moment that came just before it. Another is that we overlay memories of what we have heard before on what we are just now perceiving, comparing the two as we run the old and perceive the new simultaneously. We do not have to concern ourselves with which of these modes of comparison Aristoxenus was talking about; instead we can recognize that we use both modes together for ‘following the contents of music’. They are the complementary cognitive skills that constitute our faculty for music. Let’s consider each of them in turn.

### *Registering Relativity in Music*

When we experience music, we are tracking a chain of relativities. We register how the sound ‘that is coming to be’ relates to the one that has just ‘come to be’ in pitch, duration, volume, attack, and decay, and those relativities, whenever the new sound is different from the one before, are differences along a continuum. Musical differences, except for timbral ones, are differences of more or less. When Dorothy sings ‘Some-where’, it does not matter what pitch she sings ‘Some-’ on; it matters that she then sings ‘-where’ an octave higher. As we listen to the whole phrase, and then the whole song, the relativities in all dimensions accumulate into relationships, giving us our sense of the music’s course. And because the changes are mainly differences along a continuum, the course of a musical performance is by nature circular. What goes up in pitch also comes down, what grows louder also becomes softer, longer notes alternate with shorter ones, and the relationships that develop in the musical flux define the ending terms that the music comes around to.

Our processing of language is not comparative to the same extent. Spoken language does involve relativities of pitch, volume, and timing, and those relativities contribute to a speaker’s expression and meaning. Changes in pitch relations, for example, can change the meanings of certain words in tonal languages, and they can turn a statement

<sup>9</sup> A line of Milton famously so Latinate in word order that it is barely English nevertheless delivers its meaning unambiguously: in *Paradise Lost* V: 611–12, God the Father, speaking of his Son, decrees: ‘Him who disobeys / Me disobeys’.

<sup>10</sup> The music theory traditions of many cultures are primarily concerned with how to create good courses of musical events or how to track them.

into a question in any language. But even in tonal languages, the vast proliferation of words and meanings rests principally on phonemic distinctions drawn from a non-continuous menu of sounds: the sound of ‘f’ and the sound of ‘m’ do not occupy different places on a continuous phonemic spectrum. And once listeners sort the phonemes they are hearing into words, the relationships they are usually most occupied with are semantic and syntactic, not phonetic. At that point listeners are not concerned with comparison or relativity or circularity. In following the words of Dorothy’s song, we are not comparing each sound unit to the next (‘Some-’ to ‘-where’ or ‘Somewhere’ to ‘over’) in sound or meaning; we are separating successive sound events from each other, by drawing on our internal lexicon, and relating the words we deduce in that way to each other, by drawing on our internal grammar, until we piece together a train of thought. (Of course, when we listen to a performance of the song, we are using both our language faculty and our music faculty simultaneously – but more about that later.) And if the words are just a train of sounds to us – that is, if we don’t understand the language the words are sung in – we would say that we can’t ‘follow the contents’ of the words at all.

In music there are relativities even within a single sound event: the relative weighting of two instruments playing the same pitch simultaneously, for instance, or the interval between two voices or instruments sounding different pitches simultaneously. But those relativities in turn create the distinctiveness of a single sound event that we experience relative to the events that come before and after it, for as Aristoxenus writes, ‘it is in a process of coming to be that melody consists, as do all the other parts of music’. And in that ‘process of coming to be’ we experience an endless flow of relativities through endless juxtapositions of perception and memory. Many of these relativities register as changes – including changes from sound to silence or silence to sound – but others register as continuity or sameness. And a combination of change and continuity is also a form of relativity: a crescendo on a single or repeating note makes us register continuity of pitch in combination with change in volume.

How do our perception and our memory allow us to register each new musical sound in comparison to the previous one? This question, which Aristoxenus suggested and then left unexamined in the passage we are considering,<sup>11</sup> became a famous philosophical puzzle, retaining its fascination to the present day. Seven centuries after Aristoxenus, the Church Father Augustine of Hippo, in his *Confessions* (completed by 400 CE), undertook to elucidate perception in general by isolating it from memory. In that exercise he described the present as the sole possible object of our attention (*contuitus*), a point of time with no amplitude, ‘not divisible into even the most minute moments’, thus apparently denying perception, at least by itself, the capacity to register either continuity or change.<sup>12</sup> Cutting our perception off from

<sup>11</sup> We might suppose that Aristoxenus, as a pupil of Aristotle, at least understood the concept of memory that Aristotle articulated in the treatise known as *De memoria et reminiscencia*. On the implications of Aristotle’s ideas on memory for music, see Jessica Wiskus, ‘On Music and Memory through *Mnēmē* and *Anamnēsis*’, *Research in Phenomenology* 48 (2018), 346–64.

<sup>12</sup> Augustine of Hippo, *Confessions, Books 9–13*, ed. and trans. Carolyn J.-B. Hammond (Harvard University Press, 2016), book 11, sections 20/26–26/33, pp. 300–06.

memory, he denied us even the capacity to register ‘what is coming to be’, thus providing the severest possible challenge to what it means to experience music (an experience to which Augustine himself was, as he tells us elsewhere in the *Confessions*, extremely susceptible).

His challenge stood for millennia. Eventually it provoked a burst of fresh perspectives, coming toward the end of the nineteenth century, that golden moment when modern psychology in its formative stage was bringing scientific methods of exploration to bear on traditional philosophical questions. One such perspective in dealing with Augustine’s separation of perception from memory was to extend the Augustinian point of present time into a short duration of time, perhaps a few seconds – long enough to encompass the perception of continuity or change or both. This is what E. Robert Kelly (writing as E. R. Clay) in 1882 called the ‘specious present’:

All the notes of a bar of a song seem to the listener to be contained in the present. All the changes of place of a meteor seem to the beholder to be contained in the present. At the instant of the termination of such series, no part of the time measured by them seems to be a past.<sup>13</sup>

In 1890, adopting this idea, William James wrote:

In short, the practically cognized present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions into time ... The experience is from the outset a synthetic datum, not a simple one; and to sensible perception its elements are inseparable, although attention looking back may easily decompose the experience, and distinguish its beginning from its end.<sup>14</sup>

This concept at least allows for the capacity of an experiencing subject to perceive a continuum in time, rather than an infinity of successive Augustinian points in time. To that extent, it provides a rudimentary basis for the flowing nature of musical experience. It may even anticipate, under the rubric of *perception*, what the psychological literature more recently describes as a special kind of memory, namely *auditory sensory memory*, or *echoic memory*: the apprehension of a short sequence of auditory stimuli, retained just long enough to be processed for subsequent examination and storage before it fades.<sup>15</sup> But in joining successive events together in a common moment, it fails to differentiate memory from perception and so makes it difficult to understand how

<sup>13</sup> E. R. Clay (E. Robert Kelly), *The Alternative: A Study in Psychology* (Macmillan, 1882), 167. What Kelly called the ‘specious present’ the psychologist Daniel Stern later called the ‘present moment’, within which the ‘temporal contour’ of an experience (such as of a musical phrase) gives us the highly specific and subjective feeling that he calls a ‘vitality affect’. Daniel Stern, *The Present Moment in Psychotherapy and Everyday Life* (W. W. Norton, 2004), especially ch. 4.

<sup>14</sup> William James, *The Principles of Psychology*, 2 vols (Henry Holt, 1890; repr. Dover Publications, 1950), vol. 1, ch. 15, pp. 609, 610.

<sup>15</sup> The term *echoic* memory was introduced by Ulric Neisser in *Cognitive Psychology* (Appleton-Century-Crofts, 1967). Alan Baddeley describes the concept of *auditory sensory memory* in his standard textbook on memory, *Essentials of Human Memory* (Psychology Press, Taylor & Francis Group, 2014), 10–14.



we can register the relativity – difference or sameness – between what we have just heard and what we are now hearing.

James himself proposed a way around that difficulty. The famous passage of his *Principles of Psychology* in which he describes the ‘stream of consciousness’ goes on to analyse a kind of consciousness in which a memory is absorbed into the new perception that supplants it:

A silence may be broken by a thunder-clap, and we may be so stunned and confused for a moment by the shock as to give no instant account to ourselves of what has happened [...] what we hear when the thunder crashes is not thunder *pure*, but thunder-breaking-upon-silence-and-contrasting-with-it [...] it would be difficult to find in the actual concrete consciousness of man a feeling so limited to the present as not to have an inkling of anything that went before.<sup>16</sup>

Here, in the modest idea that an ‘inkling’ of memory lingers in a new perception, James gives us a way to understand how we register the change from the note we just heard to the one we are hearing now as an interval, or as an *accelerando*, or as a *crescendo*, that is, as a relativity and not just as a series of different sound events.

In the same era the phenomenologists Franz Brentano and Edmund Husserl developed a model that applied the same kind of thinking not just to a single act of perception, but also to the continuous action of perceiving present moments relative to past ones.<sup>17</sup> The scheme that they proposed and that Husserl named ‘temporal consciousness’ (*Zeitbewußtsein*) was that an experience of the present moment could include ‘retentions’ of earlier present-moment experiences, each successively ‘running off’, thereby creating a ‘continuity of pasts’. As Husserl wrote:

Since a new now is always entering on the scene, the now changes into a past; and as it does so, the whole running-off continuity of pasts belonging to the preceding point moves ‘downwards’ uniformly into the depths of the past.<sup>18</sup>

A whole series of retentions of ‘what has come to be’ therefore remains in the mind, though continuously in flux, to be related to ‘what is coming to be’ (also in flux). Though neither James nor Brentano nor Husserl cites Aristoxenus on the experience of listening to music,<sup>19</sup> it can be no coincidence that Husserl, in this central passage setting out (and graphing) his idea of temporal consciousness, presents as his primary example the experience of listening to a melodic phrase (as do Brentano elsewhere and

<sup>16</sup> James, *Principles*, vol. 1, ch. 9, pp. 240, 241.

<sup>17</sup> Franz Brentano, *Psychology from an Empirical Standpoint*, ed. Oskar Kraus, English ed. Linda McAlister, trans. Antos Rancurello, D.B. Terrell, and Linda McAlister (Humanities Press, 1973), 168. Original German edition, Leipzig: Duncker & Humblot, 1874. James, in a footnote to the passage cited here (p. 240) calls Brentano’s account ‘as good as anything with which I am acquainted’.

<sup>18</sup> Edmund Husserl, *A Phenomenology of the Consciousness of Internal Time*, trans. John Brough, in *The Essential Husserl: Basic Writings in Transcendental Phenomenology*, ed. Donn Welton (Indiana University Press, 1999), section 2 (‘Analysis of the Consciousness of Time’), §10, 189. Section 2 is based on Husserl’s lectures in 1905.

<sup>19</sup> Even though Aristoxenus was a noted disciple of Aristotle, it can hardly be surprising that these philosophers did not know or refer to this passage, since it occurs in a music treatise, one of the few texts by Aristoxenus to survive.

E. Robert Kelly in the passage quoted earlier).<sup>20</sup> The analyses they are all giving can apply more convincingly to music than, say, to language because it is musical experience that is more plausibly thought of as a process of tracking relativities continuously from one moment to the next. Thinking in these terms can help us, for example, understand such otherwise mysterious effects as the capacity of music to command our unbroken attention and the power of silences in music to feel like events.

But it does not help us understand other equally mysterious features of our musical experience, such as the effects of repetition. As David Huron writes, ‘with the possible exception of dance and meditation, there appears to be nothing in common human experience that is comparable to music in its repetitiveness’.<sup>21</sup> To investigate the effects of repetition and other characteristic features of music, we need to expand the time scale of our thinking from immediate juxtapositions in time to longer connections – or, in psychological terms, from echoic, or auditory sensory, memory to both working memory and long-term memory. And we also need to reorient our modelling of perception. Instead of maintaining the early phenomenologists’ framing of temporal experience as monodirectional, we need to take advantage of the methodologies of investigation and modelling developed more recently by psychologists and neuroscientists.

A crucial innovation by Ulric Neisser, a founding figure in cognitive psychology, was the modelling of perception as a cyclic process, in which schematic anticipations are continually modified by the information produced in a subject’s explorations and in which ‘the information already acquired’ determines ‘what will be picked up next’.<sup>22</sup> When we reach the cognitive neuroscience of the twenty-first century, we find further cyclic concepts of conscious processing, ones that are dependent on continuous loops of feedback and feedforward across wide areas of the brain. In the global neuronal workspace model being developed by Stanislas Dehaene, Jean-Pierre Changeux, and others, for example, those ‘recurrent loops can sustain a signal, e.g., such that it could be maintained in working memory’.<sup>23</sup> Here it is important that we not conflate the circularity of neural feedback loops (a high-speed process inaccessible to our perception) with the circularity of Neisser’s model of perception-modifying processes (a model that can be applied directly to musical experience). Still, Dehaene and colleagues are offering us a way to understand how neural feedback loops could ‘maintain a signal’ in working memory – which is to say, long enough for listening

---

<sup>20</sup> It is telling that in a century or more of phenomenological thinking about the experiencing of time, music has retained this exemplifying position. The phenomenological psychologist Daniel Stern, for example, in his study of *The Present Moment in Psychotherapy and Everyday Life*, 26, points to the musical phrase as ‘the musical analog of a present moment in ordinary life’: ‘it is felt to occur during a moment that is not instantaneous, but also not parceled out in time into sequential bits like the written notes’.

<sup>21</sup> David Huron, ‘A Psychological Approach to Musical Form: The Habituation-Fluency Theory of Repetition’, *Current Musicology*, 96 (2013), 7.

<sup>22</sup> Ulric Neisser, *Cognition and Reality: Principles and Implications of Cognitive Psychology* (W. H. Freeman, 1976), 20–24.

<sup>23</sup> George Mashour and others, ‘Conscious Processing and the Global Neuronal Workspace Hypothesis’, *Neuron*, 105 (2020), 782.

subjects not just to observe a change in music from one moment to the next but also to compare their ongoing perception of a passage of music consciously to a sustained memory of that or another passage. In order to further our understanding of the faculty for music, in other words, we need to move beyond the perception of immediate relativities to a second way of comparing perception to memory.

### *Synching Memories with Perceptions*

This second way involves simultaneously running memories of what we have experienced before – in some cases long before – in alignment with what we are experiencing in the present moment. Here we are not talking about that Jamesian trace that lingers in the temporal interface linking our immediate (echoic) memory to what we are experiencing now. Instead we are talking about a mechanism that retrieves stretches of music – from the shortest motive to entire numbers – from our short-term or long-term memory and replays them within us in sync with a new, ongoing experience that we are either producing or attending to. We’re talking about what allows you to hear a song and recognize that you have heard it before; what allows you to enjoy a jazz or gospel solo by hearing what the soloist is doing in relation to many other versions of that standard you have heard before; what makes you feel that the pattern of a song is continuing because its motives and beat patterns keep repeating; what allows you to perform a musical number from memory, unreeling your memory of it as you go; what allows you to sing along with a song after you have heard just a verse or two; what lets a child tense with excitement each time the moment for action approaches in ‘Ring around the Rosie’ or ‘Pop Goes the Weasel’.

This mechanism depends in the first place on our ability to retrieve stretches of music from memory. Scientists studying memory continue to use the traditional term *retrieval* to describe our process of bringing something stored in memory back to consciousness.<sup>24</sup> But for some time now they have understood the process of retrieving as nothing like what a canine retriever does: find, grab, and deliver. Nor do they describe a memory as a single representation, stored in one place in the mind. As neuroscientist Antonio Damasio writes, “The notion that the brain ever holds anything like an isolated “memory of the object” seems untenable.”<sup>25</sup> Instead, scientific theories of memory describe retrieval or recall as a process of reconstructing a map of an experience out of the records we made of the different kinds of impression we formed of that experience: what we heard, what we saw, when it happened, where we were, how we felt, etc. We then stored those records not in one place in the brain, but spread around the sensorimotor cortices, according to the modes of the original perceptions that were recorded. Later the brain can select these dispersed records and draw them back together as a unified memory because it has recorded ‘the coincidence of signals from neurons linked to the map’, that is, because it has identified and marked the varied

---

<sup>24</sup> Alan Baddeley, for instance, titles the ninth chapter of his *Essentials of Human Memory* ‘Retrieval’.

<sup>25</sup> Antonio Damasio, *Self Comes to Mind: Constructing the Conscious Brain* (Pantheon Books, 2010), 133.

elements of the experience as having occurred simultaneously. Damasio names this process ‘time-locked retroactivation’<sup>26</sup>; another neuroscientist, Gerald Edelman, refers to the synchronized signalling of neurons across the brain as ‘re-entry’.<sup>27</sup>

When we superimpose a musical memory – itself a composite of time-locked records – onto the new musical performance that we are perceiving, we are introducing a new element into that time-locked composite. Perception and memory feed into each other, as cognitive psychology and neuroscience tell us, in a recurrent feedback–feedforward loop. What we are perceiving guides, reaffirms, challenges, and corrects our memory, just as our memory does our perception, all through the performance. And how do we keep straight these various streams of complex, quickly changing musical information from disparate sources? How do we keep them cooperating with each other for a single minute, let alone for longer stretches, as we listen to or perform music that we remember? Part of the answer lies in the way we make long stretches of musical information manageable for our memory. As Bob Snyder explains in *Music and Memory*, we memorize music by grouping its sounds into *chunks* and developing *cues* (based on our ‘associative connections’ with each transitional moment) that signal the move from one chunk to the next.<sup>28</sup> We use the same processes to remember other temporal forms of information, too. Actors rely on them to remember their lines, just as dancers do to remember their choreography, even if the three arts offer different principles of chunking and different devices for cueing.

But another part of the answer to the question of how we coordinate memories and perceptions of music when we superimpose one on the other lies in the rhythmic nature of music. To remember any music is to reconstruct not just the sequence of its sound events, but also the proportional timing of those events within the overall flow. We would hardly be able to remember music at all, let alone coordinate our memory of it with a new performance, if it were not a progression in regulated time. As we reconstruct a musical memory in our head or in sound, we may move through it at a faster or slower speed, but in remembering most kinds of music we are engaged in reproducing the temporal proportions among its sounds. And by reconstructing its rhythmic organization, we can align that rerunning memory with the new performance we are perceiving. Our musical faculty fits us to perform this task, a task that we do not perform in the same way with language unless it is sung or chanted (i.e., rhythmically organized as music) but do perform in that way with dance moves because dance is organized rhythmically by its music. Reconstructing a musical memory while processing a new performance of the same music, aligning the two with each other without letting one obstruct or distort the other, is evidently a formidable feat, worthy of a recognition it seldom receives. But even that description does not do justice to the complexity of what we all – expert musicians and non-experts alike – accomplish as a

---

<sup>26</sup> Ibid., 141–42.

<sup>27</sup> Gerald Edelman, *Second Nature: Brain Science and Human Knowledge* (Yale University Press, 2006), 28.

<sup>28</sup> Bob Snyder, *Music and Memory: An Introduction* (MIT Press, 2000). See especially chs 4 (Short-Term and Working Memory) and 6 (Long-Term Memory).

matter of course in processing music. And if we expect psychological and neuroscientific researchers to develop convincing accounts of what goes into our ordinary musical experience cognitively, we need to spur their investigations with the fullest accounts we can muster of the ways we find perception and memory operating in us. What follows are two stabs at producing that kind of account.

The first case is a listening experience (which seems to be what Aristoxenus was principally thinking of, rather than a performing experience). Let us say you are listening to a new performance of a number you are familiar with (a performance new to you, anyway, even if it is recorded). How did you come to know that number? Was it from a lead sheet or score or an original soundtrack performance – that is, a source that can be revisited repeatedly as a touchstone? Or a conflation of memories from the ‘library’ of your own listening history? Or some combination of all these versions? Your memory of the number, then, as you listen to the new performance, will involve reconstructing not a single experience but a complex of experiences that you turn into a composite imagined performance. If the number comes from a performance tradition like jazz that emphasizes improvisation, there may not be a single standard version that you can compare the new performance of this ‘standard’ to. And even if you do have a memory of such a standard version, one or more of the other performances that you have heard (let us say a Sarah Vaughan recording of ‘Over the Rainbow’) may have recast the ‘original’ so hauntingly as to compete in your memory with the standard version (by Judy Garland in the movie), requiring you to keep negotiating what counts as your memory of the song all the while you are comparing that memory-in-negotiation to the new performance you are perceiving. And what comes of that comparison? It reveals discrepancies between the new performance and your variable memory of the number. These could be general discrepancies (the new performance for instance is taken at a tempo your memory does not prepare you for) or any number of novel alterations (new to you anyway) in an improvisatory performance. In either case, the new performance is defined for you by these discrepancies. Even if your previous memory of the number is itself a rich composite of memories, your comparative listening process defines the new performance as a particular variant of that memory, more than as simply a performance of the number.

For a second case let us consider a performing situation. You are one member of a group singing a song that you and the others have sung regularly all your lives: a religious song sung on a given day of the week or a given day or season in the year, a patriotic or team song sung at games, or an occasional song marking private rituals, such as ‘Happy Birthday’. As in the previous case, your memory of the song draws on many previous experiences, which may include striking variants of the song. But you align your own performance with the version that you decide to favour out of that composite memory. And since you are listening to your own voice as you sing, you keep aligning your performance along the way with what your memory prescribes as the song. But you are not singing alone. There are other singers, and so your perception includes another level of comparison, between what you hear from your own voice and what you hear from those other voices. Your task is to sing with the others, and that

involves aligning with them in rhythm, pitch, perhaps voice quality and other respects; and even though each singer in the group brings a different complex of memories of the song to the performance, everyone's task is the same, which means that everyone may need to negotiate between their memory and the pressure of what they perceive their fellow performers doing. As Thomas Turino writes in *Music as Social Life* about performing in bands, 'I think that what happens during a good performance is that the multiple differences among us are forgotten and we are fully focused on an activity that emphasizes our *sameness* – of time sense, of musical sensibility, of musical habits and knowledge, of patterns of thought and action, of spirit, of common goals – as well as our direct interaction'.<sup>29</sup> This sensation of solidarity does not necessarily inhibit and may even promote certain assertions of individuality. The situation and tradition may encourage individuals to strike out from the rest of the group, perhaps by embellishing expressively, or to lead the whole group in unforeseen directions. In that case you may find yourself aligning your singing contribution at one moment with a composite of your memory and your perceptions of your own and others' performances, and at the next moment you may find yourself negotiating between everyone's memories of the song (including your own) and the discrepancies one singer is creating with that composite memory.

In both these cases – the listening case and the performing case – you are making a point-by-point comparison, at an infinity of present moments, between what you remember, usually from some complex of memories, and what you now perceive, often from a complex of present sources, and you do this throughout the musical experience, keeping all these remembered and perceived performances aligned with each other from the beginning to the end of the musical experience. Even if you accomplish this action imperfectly, it is a stupendous cognitive feat. And yet it is utterly ordinary. We all do it all the time. So, how can we do it? The answer is that we are exercising this extraordinary faculty on something extraordinarily well suited to that exercise: human music. That art and the faculty for that art are 'made for each other'. That observation may suggest that they evolved symbiotically. The suggestion is well worth exploring,<sup>30</sup> but it lies outside the scope of this inquiry, which is to describe the existing relationship between the faculty and the art of music. And to do that I now turn to asking what it is about human music that gives rein to the faculty I have been describing. The answer lies in the most basic, ordinary, and widespread processes by which our music is formed, as it would have to, given that the faculty itself is shared by humans in general. These processes have of course been observed and studied in many cultural traditions. The point of reconsidering them here is to explore what it is about them that empowers our faculty for music.

---

<sup>29</sup> Thomas Turino, *Music as Social Life: The Politics of Participation* (University of Chicago Press, 2008), 18.

<sup>30</sup> Among recent studies of the evolution of music, a few that focus particularly on the evolution of the faculty, or capacity, for music are Steven Mithen, *The Singing Neanderthals: The Origins of Music, Language, Mind and Body* (Harvard University Press, 2006); Gary Tomlinson, *A Million Years of Music: The Emergence of Human Modernity* (Zone Books, 2015); and Henkjan Honing, ed., *The Origins of Musicality* (MIT Press, 2018).

## How Our Music Empowers Our Faculty for Music

### *The Flow of Time in Music*

Music measures out time. It organizes and marks time in simple arithmetic proportions. Plato describes this process in a passage of the *Philebus* in which Socrates says that one can gain real understanding of music only when one grasps, besides the system of pitches,

certain corresponding features of the performer's bodily movements, features that must, so we are told, be numerically determined and be called 'figures' (*rhythmoús*) and 'measures' (*métra*).<sup>31</sup>

This formula is broad enough in concept to be adaptable to musical practices around the world and yet precise enough to describe something remarkably consistent in those practices. Musicians tend to create 'figures' (perhaps 'motives' or 'phrases') made up of sounds in what may be highly complex successions of different time-lengths. But those different time-lengths, measured from the onset of one sound to the onset of the next,<sup>32</sup> are 'numerically determined', that is, they all relate to each other in simple ratios: 1:1, 1:2, 1:3, or 2:3. These proportional relationships, referred to as 'categorical rhythms',<sup>33</sup> are needed to allow the sounds, however rhythmically diverse they are in relation to each other, to be organized into much more even time-units. Those time-units constitute rhythmic grids (Plato's 'measures'), which are themselves 'numerically determined' by the figures, formed of the same time-lengths and divided by the same simple ratios, but providing a time-frame of regular repetitions. Examples of these grids are the *tala* in Indian traditions, the bar in Western traditions, and the *gatra* of Indonesian gamelan. In most Western music the bars are made up of beats of equal length, whereas in Middle Eastern and Balkan music the beats may be either equal in length or in 2:3 proportion to each other.<sup>34</sup> In some systems, such as the bell-patterns of certain African and African diasporan traditions or the modal rhythms of thirteenth-century European polyphony, the grid may itself be constituted of repetitions of rhythmic figures. In all, it is an elaborately mathematical system, a silent temporal grid that performers realize and project in performance through the proportional rhythms of the sounds they actually produce.

<sup>31</sup> Plato, *Philebus*, 17d, trans. R. Hackforth (Cambridge University Press, 1945), repr. in *The Collected Dialogues of Plato*, ed. Edith Hamilton and Huntington Cairns (Princeton University Press, 1961), 1093.

<sup>32</sup> Measuring musical rhythms from the onset of one sound-event to the onset of the next is a standard technique in studies of music in both humans and other species. See, for instance, Chiara De Gregorio and others, 'Categorical Rhythms in a Singing Primate', *Current Biology*, 31 (October 25, 2021), R1379. Sarah Hawkins explains why the 'p-centres' (psychological moments of occurrence) of musical notes are more closely aligned with their onsets than is the case with the p-centres of syllables in spoken words, in 'Situational Influences on Rhythmicity in Speech, Music, and Their Interaction', *Philosophical Transactions of the Royal Society B*, 369 (2014), 20130398, 3. Dancers, of course, characteristically move to the onsets of musical sounds.

<sup>33</sup> The term *categorical rhythm* for small-integer rhythmic relationships seems to be favoured especially in biologists' studies of non-human species. See, for instance, Tina Roeske and others, 'Categorical Rhythms Are Shared between Songbirds and Humans', *Current Biology*, 30 (21 September 2020), 3544–55.

<sup>34</sup> On non-equal beats, see Justin London, *Hearing in Time: Psychological Aspects of Musical Meter* (Oxford University Press, 2012), ch. 8.

This sounds so complexly mathematical and precise that you might think you would need a chronometer rather than a mere musician to achieve it. Yet Plato tells us that all these ‘numerically determined’ features correspond to ‘the performer’s bodily movements’. Here he is no doubt thinking principally of the steps and gestures of dance, as he generally does when he writes about rhythms.<sup>35</sup> Our bodies do not perform dance movements with chronometric or metronomic or click-track precision, and body-generated music does not measure time that way either, but instead with its own precision, the precision of bodily felt equivalences and proportions. And while those equivalences and proportions may be inaccurate, sometimes wildly inaccurate, by the clock, their elasticity may fall within the ‘numerically determined’ way of measuring time that participants in the given musical tradition share.

But why do we form music in such a complex time-measuring way? Plato, in the *Philebus*, does not stop to inquire. But Aristoxenus, in the passage we have already examined, points us toward the means to answer just that question. For if we need to compare ‘what is coming to be’ in music to ‘what has come to be’ in order to ‘follow the contents’ of music, we need to mark time in measured units. After all, the contents of music, as we’ve observed, lie in the sequence and flow of sounds, and without a system for measuring and tracking time in music, we wouldn’t be able to tell whether what we are hearing is flowing in a rhythm that is the same as or different from anything that came before it.

So, how does our faculty for measuring and tracking time in music work? It is one more cyclic system. As Justin London puts it, ‘in musical contexts, metric attending involves both the discovery of temporal invariants in the music and the projection of temporal invariants onto the music’.<sup>36</sup> As we listen to music, we first accumulate the sounds that we hear into measured rhythmic ‘figures’ by fitting the time-intervals we hear, from the onset of one sound to the onset of the next, into the categorical rhythms listed above. The preference for these simple proportions may ultimately derive from bodily movements such as Plato’s dance steps, as well as from oscillatory or periodic motions found more generally in our walking, breathing, autonomic processes, even our neuronal signaling.<sup>37</sup> A recent experimental study by Nori Jacoby and Josh McDermott suggests that even

<sup>35</sup> On Plato’s role in forming the Greek concept of rhythm, see Émile Benveniste, ‘The Notion of “Rhythm” in Its Linguistic Expression’, *Problems in General Linguistics*, trans. Mary Elisabeth Meek (University of Miami Press, 1971), 281–88. Originally published in French in 1951. In connecting musical rhythms to ‘the performer’s bodily movements’, Plato can be considered a forerunner of a wide range of modern concepts, from Mark Johnson’s idea that music’s meaning is embodied (see his *Meaning of the Body*) and Arnie Cox’s exploration of mimetic and metaphorical modelling of music on human bodily experience in *Music and Embodied Cognition: Listening, Moving, Feeling, and Thinking* (Indiana University Press, 2016) to psychological theories of embodied musical cognition, as in Andrea Schiavio and others, ‘Music in the Flesh: Embodied Simulation in Musical Understanding’, *Psychomusicology: Music, Mind, and Brain*, 24.4 (2014), 340–43.

<sup>36</sup> London, *Hearing in Time*, 24.

<sup>37</sup> Edward Large, ‘Resonating to Musical Rhythm: Theory and Experiment’, in *Psychology of Time*, ed. Simon Grondin (Emerald Group, 2008), 189–232, offers a ‘neural resonance theory’ of how listeners ‘experience dynamic temporal patterns, and hear musical events in relation to these patterns, because they are intrinsic to the physics of the neural systems involved in perceiving, attending, and responding to auditory stimuli’.



when we are presented with sequences of clicks in random time-intervals, we seek out and identify simple time-interval relationships between them.<sup>38</sup> But of course actual music is generated by musicians who may be conditioned in the same menu of temporal relationships as their listeners and may reinforce those relationships by accentual, tonal, timbral, and other means.

Once we locate such ‘invariant’ rhythmic relationships in the sounds we are hearing, we subject those sounds to a second fitting process: we abstract from the sounding sequence of differentiated time-units the unsounding grid that will most comfortably encompass the temporal irregularities of the sound sequence.<sup>39</sup> Then, as the music continues, this grid, constituted by the temporal relationships we ‘discover’ in the musical figures, provides the frame into which we continue to fit the temporal relationships of the figures. But though this grid is abstract and unsounding, it does not feel abstract to musicians or listeners or dancers because they feel it in their bodies as real or potential motion. That is, they feel their own bodily time-keepers drawn into sync with – entraining to – a repeating unit of time that is both measured and alive-feeling (an animating grid!), both shared with others (between one performer and the others or between listeners and performers) and apparently external to all (whether it is initiated by a single leader or in a communal process). Feeling in the grip of a rhythm that is both within us all and beyond us all produces profound effects of shared consciousness and individual well-being.<sup>40</sup> In music we entrain to the grid that we derive from the sounding rhythms, not to those rhythms themselves, and once we derive it, we sustain that grid in ourselves: we are in the groove. To be in a groove, or in general to keep music going, is to recycle a rhythmic grid; it is to reproduce the grid from our memory of what we and perhaps others have produced just before, making what we are producing and perceiving now align with that memory. And we can make this alignment happen because the grid is measured.<sup>41</sup>

---

<sup>38</sup> Nori Jacoby and Josh McDermott, ‘Integer Ratio Priors on Musical Rhythm Revealed Cross-Culturally by Iterated Reproduction’, *Current Biology*, 27 (2017), 359–70.

<sup>39</sup> The scheme proposed here agrees in large part with the ‘separable components’ that Sonja Kotz and colleagues specify as ‘underlying rhythm cognition’. What I am calling ‘abstracting the grid’, for example, they call ‘beat extraction from complex auditory patterns’. But their scheme lacks the prior stage I describe as ‘accumulating the sounds into measured rhythmic “figures”’. See S. A. Kotz and others, ‘The Evolution of Rhythm Processing’, *Trends in Cognitive Sciences*, 22.10 (October 2018), 896–97. Edward Large by contrast does account for the ‘pulse and meter’ to emerge from the complex sounding sequence of note-lengths through a process he calls *induction*; at the same time, he considers pulse and metre to be ‘intrinsic to the physics of neural oscillation. All that is then required is coupling to a rhythmic stimulus.’ Large, ‘Resonating to Musical Rhythm’, 193, 223.

<sup>40</sup> These effects have been explored in studies including William McNeill, *Keeping Together in Time: Dance and Drill in Human History* (Harvard University Press, 1995); Turino, *Music as Social Life*; Michael Hove and Jane Risen, ‘It’s All in the Timing: Interpersonal Synchrony Increases Affiliation’, *Social Cognition*, 27.6 (2009), 949–61; and Liam Cross and others, ‘How Moving Together Binds Us Together: The Social Consequences of Interpersonal Entrainment and Group Processes’, *Open Psychology*, 1 (2019), 273–302.

<sup>41</sup> The study of musical entrainment was greatly advanced by the publication of Martin Clayton and others’ ‘In Time with the Music: The Concept of Entrainment and its Significance for Ethnomusicology’, *European Meetings in Ethnomusicology*, 11 (2005), ESEM Counterpoint 1, 1–82, now on Durham Research Online: <https://dro.dur.ac.uk/8713/>. The present study benefits, for instance,

Musicians have choices beyond keeping to an established rhythmic grid throughout a performance. They can alter the grid itself: speeding it up, slowing it down, trading it in for a different model. Or they can take liberties with it. They can engage in what Charles Keil calls ‘participatory discrepancy,’ or earlier musicians called *rubato*, when one musician in a group pulls away from the grid the others are maintaining and yet never loses track of it, so that the discrepant line can always be heard and felt in relation to that rhythmic touchstone.<sup>42</sup> Or the touchstone can be recalled in silence and still make itself felt. Art Tatum, in his solo piano performances, famously rendered jazz standards with a freedom that apparently left the original rhythmic frame, as well as other features of the original song, in the dust, and yet he evidently produced all his discrepancies against a continuing memory of the model he was departing from, so that listeners who know the model, if they likewise keep drawing it from their memory as they listen to his recording, can thrill to the way Tatum measures out his discrepancies all along the way and comes out perfectly in time with the model.

A Tatum solo is no ordinary case, but it exemplifies an ordinary way that musicians and listeners use their faculty for music, just as it exemplifies the kind of pleasure that no other art than music – no less measured art – affords. The art of speech, for example, can be rhythmic in a variety of ways, but its delivery is not ‘numerically determined’ in categorical rhythms unless it is well down the road to music, that is, as chanting or rap or song.<sup>43</sup> Even the metrics of poetry deal more with counts and groupings (of syllables or stresses, for example) than with measurements of time. The pacing and stresses of spoken delivery may help us make sense of the words;<sup>44</sup> conversation can produce its own kind of entrainment, sometimes called ‘interactional synchrony’, between speakers;<sup>45</sup> and in listening to spoken language we have heard before, we may rerun the sequence of words from memory against what we are now perceiving, but that sequence will not be tied to a time-grid. An actor trying out a single line of text, even text in verse, will vary the relative lengths of syllables in the line a hundred ways, without regard to any temporal grid that would allow a listener to compare one delivery of the line to any other. But if you and I sing the same song without conveying that we are playing off, however freely, against a shared sense of the time allotments between one note and the next, it may hardly be recognized as the same song.

That said, there are certainly genres of music in many traditions that are considered temporally ‘unmeasured’. How are these genres ‘made for’ our musical faculty? In some cases the musician may perform in subtle response to a pulse (especially a very slow pulse) that listeners cannot detect.<sup>46</sup> In others, as Frank Kouwenhoven provocatively

---

from these authors’ recognition (p. 15) that ‘the motor system is not only responsible for producing a rhythm, but is also involved in the perception of rhythm’. A similar recognition of the role of memory has been less apparent so far in the work of these and other authors investigating musical entrainment.

<sup>42</sup> Charles Keil, ‘Participatory Discrepancies and the Power of Music’, *Cultural Anthropology*, 2.3 (August 1987), 275–83.

<sup>43</sup> Kotz and others, ‘Evolution of Rhythm Processing’, 900.

<sup>44</sup> *Ibid.*, 901.

<sup>45</sup> Clayton and others, ‘In Time with the Music’, 20–25.

<sup>46</sup> See Richard Widdess, ‘Involving the Performers in Transcription and Analysis: A Collaborative Approach to Dhruwad’, *Ethnomusicology*, 38.1 (1994), 59–79, especially pp. 65–68.

suggests, the musician may create a form of ‘reorganized time’ to which listeners entrain by responding to such features as internal repetitions.<sup>47</sup> These may then be cases where musicians necessarily abandon their usual strategy of measuring the flow of time and rely exclusively on the other kinds of musical measurement (such as pitch) or on attention to other kinds of organization (such as sequencing and segmenting) to engage our faculty for music; these are the modes of organization considered in the next two sections here.

The measured rhythmic proportions of music allow us to harness the human faculty for comparing across time. And they do more. Those measured proportions catch us up, giving us the feeling that the music – and we as participants in it – are moving under the spell of a mysterious propulsion, often but not necessarily a steady propulsion, through time. When we surrender to this sense of propulsion with our bodies and minds, we give the music the power to command the focus of our attention, engaging us in what may be mere sounds, sometimes for hours at a time. Ironically, because time in music is marked off by measured events, because music divides time into distinctive irregularities (rhythms) fitted into distinctive rhythmic regularities (grids), we come to feel that music is making time flow in and around us. In fact, only when we feel time being measured do we think of it as flowing at all. Music makes possible a particularly literal form of what Husserl named ‘temporal consciousness’: consciousness of time itself as the medium in which we are experiencing.<sup>48</sup>

### *Measuring Out Pitch in Music*

The sound frequencies of music are measured, just as its timings are. Measuring changes in frequency gives us a second means of comparing what is coming to be in music to what has already come to be, whether we are comparing the relative change from one moment to the next or rerunning a whole stretch of changes from memory over a new stretch of changes. But measuring frequencies, or pitches, is a much trickier business than measuring timings. In the first place, the pitch of any sound needs to be ascertainable, and for that purpose a musical sound needs to be stable; it needs to be sustained at the same frequency, more or less, at least long enough for its pitch to register in our perception. Sustained pitch not only makes musical sounds measurable in comparison to each other, but also gives music – along with some whistling sounds of natural or mechanical origin – its peculiar power to get inside human listeners and affect their nerves and emotions. Then, for us to compare pitches, one sustained pitch needs to give way to another sustained pitch (they need to be discrete), and the difference (interval) between them needs to be measurable.<sup>49</sup> But how do we measure that difference?

---

<sup>47</sup> Frank Kouwenhoven, ‘Some Remarks on Music as Reorganized Time’, Commentary in Clayton et al., ‘In Time with the Music’, 88–92.

<sup>48</sup> Victor Zuckerkandl writes: ‘Music is temporal art in the special sense that in it time reveals itself to experience’. *Sound and Symbol: Music and the External World*, trans. Willard Trask (Pantheon, 1956), 200.

<sup>49</sup> The use of discrete pitches counts as one of the most universal of musical features among those considered in Savage and others, ‘Statistical Universals’.

It is not like a temporal difference, which we can measure or count with successive steps or claps produced by repetitive motion. Nor is it like a spatial difference, which we can measure against a visual marker. (We may, though, apply spatial metaphors to sound frequency: in Western music, for instance, we speak of steps in a *scale* – from the Italian for ladder – and of high vs. low pitch.) We are measuring a difference in wave frequency that we can perceive only indirectly, since we cannot ordinarily perceive, let alone count, soundwaves. Instead we rely on a grid process, as we do in tracking rhythms, and in so doing we mark the measuring of pitch intervals as an act of memory, cultural memory. When we attend to music, we listen for intervals that we recognize from our history of listening – a history that has refined the instinctive capacities we apparently have from infancy to track directions of change and contours of melody<sup>50</sup> – and accumulate these intervals into sequences that we then seek to fit into a grid of intervals familiar to us. The grid is composed of a set of discrete steps. In Western music such a set is called a scale or mode. Concepts such as *maqam* in Middle Eastern music or *raga* in Indian music include a set of steps, but also specify more about the expected shaping of melodies from those steps.

The relationships of frequencies along this grid are to some extent determined in simple integers, just as rhythmic relationships are, but in this case those integer ratios are not perceptible as such: we do not hear or feel that pitches an octave apart have a 2:1 relationship in frequency. And yet it is ‘statistically universal’ that the frequency grids of musical systems are devised within the interval of the octave. No doubt human nerves are prone to respond to the acoustics of the octave, that is, to the way the frequency of one pitch fits into the frequency of the pitch an octave lower. But no human culture or tradition employs a frequency grid that is entirely reliant on acoustic properties in simple integer relationships. As Alexander Ellis, one of the earliest scholars to attempt a global survey of scales and the one who introduced the system of cents to measure intervals, wrote in 1885: ‘The musical scale is not one, not “natural”, not even founded necessarily on the laws of the constitution of musical sound so beautifully worked out by Helmholtz, but very diverse, very artificial, and very capricious.’<sup>51</sup>

That is to say that human scales, employing some combination of simple-integer intervals and more ‘capriciously’ devised ones to fill the octave, tend overwhelmingly to be constituted of unequal steps. This kind of grid allows musicians to organize musical progressions in a way that they and their listeners can track and keep track of (i.e., remember), not necessarily in an end-oriented way, so that they can use their music faculty to overlay a remembered pattern of pitches, in its remembered

<sup>50</sup> Sandra Trehub, ‘Human Processing Predispositions and Musical Universals’, in *The Origins of Music*, ed. Nils Wallin and others (MIT Press, 2000), 427–48, esp. pp. 428–31.

<sup>51</sup> Alexander Ellis, ‘On the Musical Scales of Various Nations’, *Journal of the Society of Arts*, 33 (1885), 526. Reprinted in Kay Kaufman Shelemay, ed., *Garland Library of Readings in Ethnomusicology* 7 (Garland Publishing, 1990), 1–43. This observation places Ellis in Aristoxenus’ rather than Pythagoras’ line of scale theorists.

rhythm, over what is coming to be.<sup>52</sup> Scales constituted of equal steps, by contrast, are disorienting.<sup>53</sup> Furthermore, the principle of uneven steps allows the musicians of any musical culture or tradition to choose among different kinds of scales, even to switch among them within a single musical number or song, thereby enriching the expressive possibilities of their repertoires. In many kinds of music, ornamental gestures such as glissandos and scoops may smear the scale briefly, making in effect excursions away from its path, but they generally start from and end up on the steps of the scale. Only certain percussion sounds – those that have either no detectable pitch or too rich a mixture of pitches – really evade the scale altogether. They are generally vital to the music's rhythmic propulsion, leaving it to other sound sources to deal with the realm of pitch. But music may be made up entirely of such percussion sounds, in which case our faculty for music focuses entirely on rhythmic and other non-pitch patterns, just as in unmeasured music it focuses entirely on pitch and other non-rhythmic patterns.

Like the rhythmic grid, the scalar or pitch grid of any music emerges from the sounds we hear. That is, we abstract it from the process of comparing the pitches that are coming to be to those that have come to be. And as with the rhythmic grid, the effect of this abstracting is not abstract. Through experience of listening to music that uses a particular scalar system, we develop a sense of the relationship between any given tone and its scale, and then when we follow the course of a given song, our faculty for comparing perception to memory orients us in the course of that melody, along with any drones or counter-melodies or harmonies that come with it. The orienting power we derive from the combined scalar and rhythmic grids in ordinary musical listening and performing activities gives us an astonishing capacity to detect minute alterations in choice of pitch from one iteration of a phrase or song to another.<sup>54</sup>

Let us say for instance that as someone reasonably familiar with the grids of Western classical music, you are listening to Franz Schubert's song 'Du bist die Ruh'. The strophe beginning with the line 'Dies Augenzelt, von deinem Glanz' takes about half a minute to perform, during which time we hear about a hundred separate notes from the singer and pianist (Example 1).

After a short interlude, we then hear the same strophe again, with the same words. This time there is a tiny change in the vocal line. Near the beginning, not even at the climax of the line, one repeated pitch, on the word 'deinem', is slightly altered (Example 2).

---

<sup>52</sup> Even infants evidently process music more effectively when it is made of uneven-step rather than even-step scales. See Sandra Trehub and others, 'Infants' and Adults' Perception of Scale Structure', *Journal of Experimental Psychology: Human Perception and Performance*, 25.4 (1999), 965–75.

<sup>53</sup> It is precisely for disorienting effects that musicians sometimes adopt equal-step scales (e.g., the chromatic and whole-tone scales in Western music).

<sup>54</sup> By ingenious experimental design Sandra Trehub and colleagues produced evidence that six- to eight-month-old infants can respond differentially to a six-note melody when it is repeated with a single pitch significantly changed. See Sandra Trehub and others, 'Infants' Perception of Melodies: Changes in a Single Tone', *Infant Behavior and Development*, 8 (1985), 213–23.

Example 1 Franz Schubert, 'Du bist die Ruh', D. 776 (published 1826), words by Friedrich Rückert, bb. 54–67.

Dies Au - gen - zelt, von dei - nem Glanz al - lein er - hellt, o - füll es ganz, o - füll es ganz!

Example 2 Schubert, 'Du bist die Ruh', bb. 68–82.

Dies Au - gen - zelt, von dei - nem Glanz al - lein er - hellt, o - füll es ganz, o - füll es ganz!

Even if you do not understand the language being sung, you are unlikely to miss this alteration and may even find it hair-raising.<sup>55</sup> But how do you even detect that difference in a single pitch among so many, let alone feel its expressive force?

<sup>55</sup> Listeners frequently find music at certain moments hair-raising, or chilling, or tear-inducing, and studies have explored what kinds of events in music set off such physical responses. See Donald Hodges, 'Bodily Responses', *Oxford Handbook of Music Psychology*, ed. Hallam and others, 183–96.

You follow the singer's melodic line in the first strophe by comparing its movement to the rising form of its scale. The particular scale here comes as a surprise at this point in the song, but you are carried along by the perfect match of the melody to the ordering of that scale. Then at the repetition of the strophe, you make a more complex comparison, using the scale to align the present melodic line with what you heard just before (as you do when you use an ongoing metre to align a rhythm that is coming to be with one that you remember). This orienting force of the scale makes the change of pitch on 'deinem' not just noticeable, but potentially wrenching. You feel the move this time from the *c* of 'von' to the *f* of 'deinem' not just as an alteration to what you heard before, but also as a disruption of the scalar movement that the earlier version had embodied so regularly, as well as a comparative enrichment of the harmony, itself a product of the scale. The disruption feels resolved once the line is absorbed back into the progression of the scale.

This use of the faculty for music has no real equivalent in our experience with spoken language. Control of pitch is of course tremendously important in spoken language, and in several ways. But pitch is more fluid, less discrete in most spoken language (with exceptions such as street hawking calls and preaching styles that verge on song) than in most music. That fluidity in itself makes it harder to measure pitch intervals in speech. In listening to spoken language, we track contours of pitch, glean expression and meaning (e.g., distinguishing question from statement in some cases); and in conversational exchanges speakers may match the contours of each other's remarks or may dovetail a response with a question by means of pitch.<sup>56</sup> In tonal languages the very identity of a word may depend on the speaker's use of pitch. But though speakers of the same tonal language may speak a word with the required rise or fall, they will not use the same pitch interval, as they would in singing; instead, they will typically adjust their intervals to suit their individual speaking ranges. As the linguist Robert Ladd writes, 'linguistic equivalence of pitch between speakers with different ranges is not based on anything like a musical scale'.<sup>57</sup> All of these linguistic cases involve sensitivity to pitch change, but none of them entails what we do in producing and listening to music: track the organization of a sequence or phrase of pitches point-to-point along a measured grid, or scale.

We do not remember the contour of spoken language in that way, and we do not replicate it in that way. We cannot tune our speaking voices with minute precision – what would that even mean? – as some musicians in many traditions devote prodigious

---

<sup>56</sup> Jan Gorisch and others, 'Pitch Contour Matching and Interactional Alignment across Turns: An Acoustic Investigation', *Language and Speech*, 55.1 (2012), 57–76; Juan Pablo Robledo and others, 'Pitch-Interval Analysis of "Periodic" and "Aperiodic" Question+Answer Pairs', *Speech Prosody* 2016, 1071–75. <https://eprints.whiterose.ac.uk/96973/>.

<sup>57</sup> D. Robert Ladd, *Intonational Phonology*, 2nd edn (Cambridge University Press, 2008), 196. Likewise Ray Jackendoff writes that 'there is no convincing analogue in language to the music use of pitch space, despite their making use of the same motor capacities in the vocal tract'. 'Parallels and Nonparallels between Language and Music', *Music Perception*, 26.3 (2009), 200. For further comparison of language to music as a sound system, see Aniruddh Patel, *Music, Language and the Brain* (Oxford University Press, 2008), especially ch. 2.3, 'Linguistic Sound Systems'.

efforts to tuning their singing voices and instruments, and therefore spoken language cannot induce the exquisite pleasures of intonation that finely tuned music can produce. Spoken words are never out of tune, nor do we say you have made a mistake (let alone slap your wrist with a ruler) if you speak them at the wrong pitch. And speakers cannot bend their pitch, as musicians in folk, classical, and popular traditions on every continent do. That practice makes an effect on listeners only because they are comparing the bent pitch to – measuring it against – the unperformed pitch that the scale leads them to expect. That is, they are feeling the slight but acute difference between what they perceive coming to be and what their memory tells them should come to be. Finally, we do not entrain to pitches in speech the way we do to pitches in music, deriving sometimes overwhelming experiences of social connectedness from singing or playing in unison or in harmony. Even the simplest pitch entrainment, as when the partisans of one team in a sports stadium come together singing a single prolonged pitch, without rhythm, without words, and without movement to a different pitch, can create a powerful feeling of oneness for their side and a powerful disheartening effect on the other team. Pitch entrainment has not been researched as intensively as rhythmic entrainment, but the study of the faculty for music, with its reliance on both measured time and sustained, discrete pitch suggests that the two ensuing forms of entrainment deserve the same level of recognition and study.

### *Sequencing in Music*

Temporal and frequency grids provide frames for recalling, performing, hearing, and comprehending musical progressions. They act like the walls of a canal, both restricting the music's path and directing its motion forward. To perform music or listen to it is to feel yourself moving, or being moved, along its particular path through time. This feeling of movement comes in part from what I have described earlier as registering an endless chain of relativities between what is sounding one moment and what is sounding the next. And in part it comes from tracking, on a longer timescale, whole chains of sounds against the chains of virtual sounds that we replay from our memory. And how do we do that? How does anyone hold enough of a record of music in their consciousness to rerun a phrase of it, let alone a whole song or dance, in memory or in performance or to follow it, as Aristoxenus says, and comprehend it? After all, even if a given musical number works with a limited range of pitches and durations, their possible combinations are infinite. Musicians may hold the whole of a musical number, at least schematically, in their long-term memory, but as Husserl's model of ever-fading retentions illuminates, that whole memory cannot be conscious at any one time. Roger Chaffin and colleagues describe the process of producing music from memory as 'associative chaining': 'what you are playing reminds you of what comes next'.<sup>58</sup> But even using this process, we cannot recall a long musical sequence as an undifferentiated whole.

---

<sup>58</sup> Roger Chaffin and others, 'Performing from Memory', *Oxford Handbook of Music Psychology*, ed. Hallam and others, 560.



Instead we treat music, in performing or listening, as a series of segments. We do not arbitrarily divide music into segments as we engage with it, the way people divide long lists of digits into segments in order to remember them.<sup>59</sup> No, in music the segmenting is built into the structure, and performers and listeners pick up on it. ‘Memorability’, Bob Snyder tells us, ‘is related to how “chunkable” a sequence is, which will depend on the amount of repetition and on boundaries formed by discontinuities in a sequence.’<sup>60</sup> It helps that musical sequences are often spun from segments, or chunks, of more or less comparable duration. Phrases, for example, are segments of about the length of time that a singer can sing before needing a breath. The completion of that segment gives the singer pause to breathe and so in a literal sense allows the music – and the listeners’ apprehension of it – to continue. But it also signals to listeners the rounding off of a unit of attention, which is to say an object of comparison, and that comparing allows listeners to incorporate the completed phrase into their sense of the ongoing progression of the music. For that reason instrumental music, in particular, is often shaped into what we apprehend as phrases even when there is no breath or rest between those phrases.

Repetition is the ultimate segmenting device in music. That is, the recognition that something is repeating teaches us that one segment has ended and another begun. Elizabeth Hellmuth Margulis, whose important study *On Repeat: How Music Plays the Mind* examines the unique pervasiveness of repetition at all levels of musical experience, writes: ‘for music in a novel style, repetition is often a listener’s first way into what counts as a unit – what should be grouped together and treated as an entity’.<sup>61</sup> Repetition is a form of circularity in which the end of a segment cues the return to its beginning. This cue provides a crucial reassurance to both performers and listeners, since otherwise the process of ‘chain association’, or relativity-registering, may well come to a pause at that moment, leaving all parties unsure of what comes next.

Our temporal and frequency grids are themselves systems of repetition, the fundamental patterns of repeating time-units and pitch-units (the steps of two or more sizes repeated within a scale) that play abstractly in our heads and bodies, allowing us to apprehend the segments of actual sound content that play on and with them. And those segments are constituted, to a remarkable extent, by repetitions, whether total or varied or partial, sometimes at just the outset of a segment, sometimes at just its close, sometimes in just the rhythmic domain, sometimes (more rarely) in just the sequence of pitches, sometimes in both but with changes in volume or timbre or other domains. In a stanzaic song, for instance, whether it is an African American blues or a European folk ballad or an Arabic *muwashshah*, it is not simply the melody of the stanza as a whole that is repeated; that melody may also contain within it a host of repeating rhythmic motives and melodic motives, to say nothing of verbal repetitions. It is this pervasive repetition of segments at many levels, together with the temporal and scalar grids, that

---

<sup>59</sup> Baddeley, *Essentials*, 21–25.

<sup>60</sup> Bob Snyder, ‘Memory for Music’, *Oxford Handbook of Music Psychology*, ed. Hallam and others, 168.

<sup>61</sup> Elizabeth Hellmuth Margulis, *On Repeat: How Music Plays the Mind* (Oxford University Press, 2014), 23.

allows pairs of South Indian violinists or of jazz soloists or small groups in other musical traditions to perform staggeringly long and intricate melodies in flawless unison. And in any music, repetition in content from one segment to another allows whatever is different to draw a listener's attention, as that one changed pitch in our Schubert song does. Within the domain of musical sounds, everything comes down to the question of whether we are hearing the same thing as before or not.

The many kinds and forms of repetition in music allow us, in music we are hearing for the first time, not only to compare what we are hearing to what we have already heard, but also to anticipate what is to come. The immediate anticipations that Husserl describes as 'protentions' arise from a continuous processing of experienced presents that can create a sense of what the next experienced present will or at least might be. That sense surely depends on our comparative process allowing us to detect a pattern in those experienced moments, a pattern of continuation or progression or variation or alteration, any of which relies on some underlying repetition.<sup>62</sup> Anticipation can also arise from a different form of repetition, the repetition of conventional patterns or formulas across one's whole experience. Every new time we hear such a well-known pattern beginning, we anticipate its usual completion, whether it is a dominant chord 'resolving' to its tonic or the flow of gamelan notes leading to a gong stroke. As David Huron has analysed masterfully in *Sweet Anticipation: Music and the Psychology of Expectation*,<sup>63</sup> both the fulfilment of familiarized patterns and the frustration of them give music its power to manipulate our emotions.

Then too, the familiarity of a particular song we have heard many times is a function of repetition, and that familiarity allows us to anticipate our own response to certain moments of manipulated expectation and therefore to respond to what is unsurprising with what is often more intensity than to a comparable effect in a song we don't know. Neuroscientific research by Valorie Salimpoor and colleagues at McGill University has revealed the connection of anticipation in hearing familiar music to the discharge of neurotransmitters such as dopamine in the brain. In doing so, this research illuminates the human obsession with hearing or performing the same music over and over again (while our appetite for hearing or reciting the same prose or speech repeatedly is much weaker).<sup>64</sup>

Segmenting, of course, is as necessary a part of speech as it is of music, since people need to take regular breaths when they speak as well as when they sing. But speech can be effective without continuing from one segment, one phrase, to another: 'It's raining.' Music cannot. Without comparing what we are hearing with what we have

<sup>62</sup> As Nicolas Ruwet writes, 'if it is true that variation is the soul of all music [...] it is no less true that to say *variation* is to say *repetition*: there can only be variation on a given level, whatever that may be, if there is at the same time repetition on another level'. *Langage, Musique, Poésie* (Editions du Seuil, 1972), 136.

<sup>63</sup> David Huron, *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, 2006).

<sup>64</sup> V. Salimpoor and others, 'Anatomically Distinct Dopamine Release during Anticipation and Experience of Peak Emotion to Music', *Nature Neuroscience*, 14.2 (February 2011), 257–64. For a survey of research on the roles of neurotransmitters in musical experiences, see Yuko Koshimori, 'Neurochemical Responses to Music', in *Oxford Handbook of Music and the Brain*, ed. Donald Hodges and Michael Thaut (Oxford University Press, 2019), ch. 14.

already heard, as Aristoxenus tells us, we could not comprehend music. Comprehension of speech also requires the ability to segment what we are hearing, especially to discern within the flow of sound what constitutes the end of one word and the beginning of the next. Anyone who has struggled to understand speech in a new language can confirm how crucial that skill is to comprehension. But the process of speech segmentation is not performed for the purpose of comparing so much as it is of identifying.

Likewise, speech is nothing like as repetitive as music, except when it is turning into music, as in ‘This little piggy’ or ‘government of the people, by the people, for the people’. Repetition is music’s game – Margulis calls it ‘a fundamental characteristic of what we experience as music’<sup>65</sup> – and we can say so because other arts play different games. In the visual arts, for instance, Aristotle supposed that much of the pleasure in viewing an image lies in discovering relationships across domains, between the image and what the image imitates. ‘The reason why we enjoy seeing likenesses’, he writes in the *Poetics*, ‘is that as we look, we learn and infer what each is; for instance, “that is so and so”’.<sup>66</sup> Verbal arts likewise generate responses through recognition of objects in another domain – though the objects in that case may be thoughts or feelings as well as the experiences of our senses. Compared with the visual and verbal arts, music has a limited capacity for directly representing or identifying what is outside its own realm; even in imitating sounds, music is most effective when those sounds themselves have musical qualities, as when Siberian or North American throat singers render the sounds of birds and other animals. Nevertheless, music is undoubtedly powerful at stimulating motions real, implicit, or imaginary; at evoking memories, images, and emotions; and at controlling the path of one’s consciousness. Its power is inherently imprecise in its object, compared with what language or visual art can do, though cultural tradition or individual experience may focus our associations. How music creates associations at all with so little capacity for directly identifying an external object is an age-old question. But any attempt to understand how music makes us feel connected to something different from itself needs to take into account the one way in which music is extraordinarily imitative: through repetition, including variation, music endlessly imitates itself.

That self-imitation leads to some of the most distinctive effects of musical experience. As remembered and ongoing sound segments overlap and repeat within the confining canal walls of musical time and pitch frames, music becomes an experience of circularity. That circularity in turn gives us the feeling, noted earlier, that the music is a living spirit within us and we are living within the music. Whatever images or associations or narratives we read into the music feel as if they are arising within us – where else could they come from? – and whatever solidarity we feel with those around us through entrainment likewise seems to derive from bringing those people into the inner world of our feelings. What might be our default state of consciousness – our

---

<sup>65</sup> Margulis, *On Repeat*, 5.

<sup>66</sup> Aristotle, *Poetics*, 1448b, trans. W. H. Fyfe in *Aristotle in 23 Volumes*, vol. 23 (Harvard University Press, 1932).

sense of being located in the here and now – loses its grip on us and we are transported, sometimes a bit distractedly, sometimes quite drastically (as in trance experiences), to other states: to realms of memory (such as the memory of when and where we first heard that song<sup>67</sup>), of reverie or fantasy (such as the feeling of being transported to ‘somewhere else’<sup>68</sup>), or even to a different identity.<sup>69</sup> These universal shared experiences of music all spring from our capacity to surrender ourselves to the segmenting, recycling, recalling, overlapping, and comparing of sound patterns measured in time and pitch.

### Converging Faculties?

Throughout the previous sections my focus has been how our experiencing of music differs from our experiencing of language. I have done that to isolate how music is shaped to and by a cognitive faculty specific to music. So far I have not described the faculty that we might presume language is likewise shaped to and by: the faculty for language. But though music and language are in many ways different, we endlessly draw them together because we so commonly produce music with our voices, and when we do, we so commonly produce it with words, in a single stream or utterance. That raises the question of how the respective cognitive capacities – our faculties for music and for language – work together when we are engaged in song. And to address that question, I now need to consider the nature of the language faculty.

Among theoretical linguists and researchers in related fields there has been lively debate for at least a couple of decades about how to describe our cognitive faculty for language, in terms similar to those I have been using to describe the faculty for music. They have in mind our faculty for forming and apprehending linguistic utterances, a faculty whose operation accounts for the range of forms that human languages take. The most prominent models of this faculty are one proposed and subsequently developed by Noam Chomsky and associates and another proposed and developed by Ray Jackendoff and associates.<sup>70</sup> Both models focus on our ability to connect

<sup>67</sup> John Booth Davies described what he called the ‘Darling, They’re Playing Our Tune’ phenomenon as a distinctive trait of musical experience in *The Psychology of Music* (Stanford University Press, 1978).

<sup>68</sup> See Alf Gabrielsson, *Strong Experiences with Music: Music Is Much More Than Just Music*, trans. Rod Bradbury (Oxford University Press, 2011); Ruth Herbert, *Everyday Music Listening: Absorption, Dissociation, and Trancing* (Routledge, 2012).

<sup>69</sup> See Gilbert Rouget, *Music and Trance: A Theory of the Relations between Music and Possession*, trans. author, rev. Brunhilde Biebuyck (University of Chicago Press, 1985; original French edn 1980), the classic study of the subject. Judith Becker’s study, *Deep Listeners: Music, Emotion, and Trancing* (Indiana University Press, 2004), brings more recent neuroscientific findings to bear on the subject.

<sup>70</sup> The model by Chomsky and associates appears in Marc Hauser, Noam Chomsky, and W. Tecumseh Fitch, ‘The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?’ *Science*, 298 (2002), 1569–79, updated in Robert C. Berwick and Noam Chomsky, ‘The Biolinguistic Program: The Current State of Its Development’, *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, ed. Anna-Maria Di Sciullo and Cedric Boeckx (Oxford University Press, 2011), 19–41. The Jackendoff model appears in Ray Jackendoff, *Foundations of Language: Brain, Meaning, Grammar, Evolution* (Oxford University Press, 2002) and Jackendoff,

individual words (sounds associated with meanings) into thought-bearing utterances such as sentences through syntactical operations. These operations notably include what is known as recursion: the embedding of structures within structures (phrases within clauses, for instance, or clauses within sentences).<sup>71</sup> Our language faculty, that is, creates or apprehends an utterance not strictly as a succession, or concatenation, of words, but as a set of relationships of meaning derived from syntactic dependencies created or discerned within and across sets of words. This linguistic faculty, then, differs crucially from our musical faculty, relying as that does on the tracking of measured relativities as one sound event succeeds another and on the comparison of ongoing successions of those events to those that preceded it. The question for our cognition of song, then, is: how can we employ these two faculties – two such different methods of processing a flow of sounds in time – simultaneously, without each tripping up the other?

To look at the question from a somewhat different angle, the two faculties make use of memory in very different ways. Jackendoff describes words as ‘long-term memory linkings of structured sound, syntactic features, and structured meaning’.<sup>72</sup> To generate or apprehend a verbal utterance, that is, requires using one’s long-term memory to draw individual words, with all their linkages, from the vast lexicon that a speaker of a language knows – not an infinite set, to be sure, but many thousands (much larger than the set of pitches or time-values drawn on to generate or apprehend a piece of music).<sup>73</sup> It further requires using long-term memory to call on the complex rules of syntax, morphology, and word order that are used to form words into a thought. But the language faculty, as we have seen, does not have such a need as the music faculty has to draw from both short-term and long-term memory the measured relativities of timing and pitch of a sequence of sounds. For the language faculty, in other words, memory is largely a sourcing resource; for the music faculty it is largely a reproducing one.

When speech and music are united in song, the words are tucked into the very structures of music that allow a person to regenerate a stream of musical sound from memory: measured, proportional rhythms; measured, proportional pitches; and segmenting and repetition of all sorts and at all levels. Anyone who can sing a song from memory knows how closely words and melody are tied together in memory and how hard it can be to recall and speak those words by themselves when we have learned them

---

‘What Is the Human Language Faculty? Two Views’, *Language*, 87.3 (2011), 586–624. See also the exchange between the two parties in Steven Pinker and Ray Jackendoff, ‘The Faculty of Language: What’s Special about It?’ *Cognition*, 95 (2005), 201–36; W. Tecumseh Fitch, Marc Hauser, and Noam Chomsky, ‘The Evolution of the Language Faculty: Clarifications and Implications’, *Cognition*, 97 (2005), 179–210; Ray Jackendoff and Steven Pinker, ‘The Nature of the Language Faculty and Its Implications for Evolution of Language’, *Cognition*, 97 (2005), 211–25.

<sup>71</sup> Hauser, Chomsky, and Fitch, in ‘The Faculty of Language’, 1571, claim that recursion is a ‘core property’ of the language faculty and ‘appears’ possibly ‘to lack any analog in [...] other domains as well’. Jackendoff, in ‘What Is the Human Language Faculty’, 591–99, claims that recursion is part of a general human capacity that includes music.

<sup>72</sup> Jackendoff, ‘What Is the Human Language Faculty’, 599. On p. 587, n. 3, he notes that Hauser and others, in ‘The Faculty of Language’, p. 1571, ‘specifically exclude memory from [the language faculty, in the broad sense], for reasons unclear to me’.

<sup>73</sup> The vast lexicon of words in any language is of course built from a much smaller set of phonemes.

in union with their melody. For that matter, people with various kinds of dementia who have lost or largely lost their ability to speak are often able to sing along with someone else in a song they have known for a long time, still able to call up the words as they pour out the melody.<sup>74</sup> Evidently in these people a largely intact faculty for music delivers the words to them along with the melody through its process of point-by-point comparison in sound, while the impairment of their faculty for language prevents them from expressing themselves coherently whenever they try to generate spoken sentences. That is not to say that the faculty for language shuts down when we are dealing with songs. So, how do the faculties of language and music accommodate and influence each other in this utterly commonplace collaboration?

One answer is simply that ‘something’s gotta give’. As the literary scholar Mark Booth writes, in *The Experience of Song*:

The ways in which song words are subject to the pressure of their music are subtle and fascinating. They are reinforced, accented, blurred, belied, inspired to new meaning, in a continual interplay. In that interplay there is a constant tug against the resolution of the words to carry out their own business. The words must have an internal discipline to maintain their integrity in their cooperation and in their competition with the music.<sup>75</sup>

Booth is telling us about the interplay of the two faculties here, even if his language slyly personifies words and music themselves as the protagonists of the interplay, because his subject is after all our ‘experience of song’. The cooperation and the competition that he describes are real issues for our two faculties, because our capacity as a whole has its limits. The simultaneous comparison of overlapping sequences that the faculty for music performs is such a complex and absorbing process that engaging in it may not leave us as much capacity for complex thought in complex syntactic form as we have when our faculty for language is dealing with plain speech or writing. As a result, when we are listening to song, we may lower the demands we ordinarily make on ourselves to sort out difficult, even moderately obscure or complex, meanings in words.

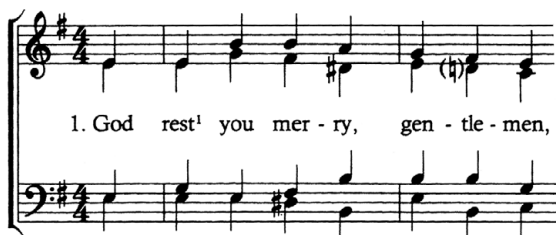
Take the case of a song that millions of people in the English-speaking world have sung and listened to endlessly (if seasonally) through their lives: the Christmas carol ‘God Rest You Merry, Gentlemen’ (Example 3). How many of us have ever asked ourselves what that opening phrase means? The carol is as much as five centuries old and has been treated to any number of variants over those centuries,<sup>76</sup> some of them seeking to compensate for the fact that the verb ‘rest’ has long since lost its transitive sense, which originally made the line mean ‘God keep you merry, gentlemen’. As a result, most of us assume that we are putatively singing to some ‘merry gentlemen’, ignoring the location of the comma, which stares at us reproachfully if we happen to be singing from a reputable source like the *New Oxford Book of Carols*, and then fail to ask ourselves what even ‘God rest you, merry gentlemen’ might mean. We could of course

<sup>74</sup> See Lola Cuddy and others, ‘Memory for Melodies and Lyrics in Alzheimer’s Disease’, *Music Perception: An Interdisciplinary Journal*, 29.5 (2012), 479–91.

<sup>75</sup> Mark C. Booth, *The Experience of Song* (Yale University Press, 1981), 7–8.

<sup>76</sup> Charles Dickens, in *A Christmas Carol* (1843), cited the song as ‘God bless you, merry Gentlemen’.

Example 3 'God Rest You Merry, Gentlemen', traditional Christmas carol, arranged by John Stainer in *Christmas Carols New and Old* (London, 1871) and republished in *The New Oxford Book of Carols*, ed. Hugh Keyte, Andrew Parrott, and Clifford Bartlett (Oxford University Press, 1992), p. 522, bb. 1–2.



look up the very fine entry on this carol in Wikipedia,<sup>77</sup> but mostly we do not do that, not because we are lazy, but because we do not expect songs to make sense the way we expect prose or poems to.

And if we did, we would be up against the music itself in this case. The opening phrase of this melody marches in irresistible rhythm down the scale from the dominant on the first syllable of 'merry' to the tonic on the last syllable of 'gentlemen', making it virtually impossible for singers to disconnect 'merry' from 'gentlemen' and thereby associate it with 'God rest you', or for listeners to hear that articulation. 'The pressure of their music' undoes – Booth might say 'belies' – the 'resolution of the words to carry out their own business'. In this case, then, the music seems to draw on our faculty for music in a way that impedes the ordinary operation of our faculty for language.

But the faculties for language and music may also collaborate in song to induce levels of consciousness in us that neither one could give us on its own. Consider how melody makes words, phrases, or whole speeches stick in our memory, sometimes calling them up unbidden when something about our situation makes room for the thought and sentiment that those words and that melody embody for us. Let us suppose, for instance, that in thinking about a certain unhappy stage in a love life, our own or someone else's, we find ourselves recalling the opening line of a song that we may have heard many times, the Rodgers and Hart standard 'I Didn't Know What Time It Was'. Here it is the faculty for language that leads the way, connecting the state of emotional confusion we are thinking about to a remembered phrase that refers metaphorically to that state as 'not knowing what time it is'. But it is the song that comes back to us, not just its words, and in fact the words are so trite that we would hardly remember them if they were not, as we noticed about songs earlier, tucked into musical structures that are shaped by the music faculty: measured rhythms and pitches, phrase segments and repetitions. And those are the structures that allow us to regenerate the sounds of a song, including its words, from memory.

When the opening phrase of that familiar Rodgers and Hart song comes back to us, we are unlikely to experience it as gloomy, even though the words are about an unhappy state of mind. But then, the song as a whole is not gloomy. The very next

<sup>77</sup> Wikipedia, 'God Rest You Merry, Gentlemen', consulted 1 September 2022.

line reports the subject's escape from unhappiness: 'Then I met you'. But if we know the song already, we do not wait for the second phrase to feel the gloom lift. We feel it already when the first phrase comes to our mind. And if a favourite singer in a nightclub launches into this song, the audience may recognize it with delight from the first phrase and burst into applause. The secret is that the opening words and music bring the whole song, including its emotional trajectory, back to us from our previous experiences of it. Antonio Damasio, thinking of his own experiences listening to the music of Bach, writes of this kind of recall as drawing on a collection of past episodes:

In certain situations, the number of summoned episodes can be very high, a true flood of memories suffused with the emotions and feelings that first went with them.<sup>78</sup>

To those who know the song, in other words, the approaching solution to the unhappy subject's problem is already present in the opening phrase, suffused with the listener's joy in contemplating that solution, even though the opening words recall only the problem. From the perspective of the study of neurotransmitters by Salimpoor and others discussed earlier,<sup>79</sup> this anticipation of the second phrase may trigger the release of dopamine in listeners as they are listening to the first phrase. Or, adapting the vocabulary of phenomenological analysis to the experience of hearing a remembered song, we might describe the effect here as a retention of a protention (a projection of the line to come from our memory of the whole song) or as a saddle-back of time from which we are viewing the present situation as the future of the past one. But however we describe it, the effect is intensified by the simplest of musical effects: Rodgers has given the second line, 'Then I met you', almost an identical setting to the opening 'I didn't know what time it was'.<sup>80</sup> In these two sung phrases our faculty for language, with its aptitude for apprehending narrative, allows us, even if we are hearing the song for the first time, to look back in time from a happy present to a confused past. Meanwhile our faculty for music, with its aptitude for overlapping comparison, allows us, through this musical echo, to feel the earlier pain as a continuing point of comparison in the present joy. And, if we know the song already, our faculty for music allows us to feel that poignant joining of past and present in a rush at the moment when we hear the song's opening in a new performance.

Our faculties for language and music make such different demands on our attention as we listen to a song that they may interfere with each other, but they may also, in a case like this, combine their two kinds of consciousness to lead us into levels of experience we could not attain without engaging both faculties. And if studying the operations of the two faculties in an apparently simple case like this song opening can illuminate our listening experience of it, then we can use the same analytical framework to investigate the roles of memory, the shifting of our focus between words and music, and the surges of emotion that characterize our experience of a wide range of songs.

<sup>78</sup> Damasio, *Self Comes to Mind*, 211.

<sup>79</sup> See Salimpoor and others, note 64 above.

<sup>80</sup> These phrases can be heard at 0:53 sung by Ella Fitzgerald in *The Rodgers and Hart Song Book* <[https://www.youtube.com/watch?v=τ78xp\\_BHXIA](https://www.youtube.com/watch?v=τ78xp_BHXIA)> (accessed 9 May 2024).



## And How About Our Fellow Species?

What can our understanding of the human faculty for music tell us about the musical nature of other species? It has been common since ancient times for humans to extend concepts such as *song* and *music* to cover sounds produced by other species. The Roman philosopher poet Lucretius even traced the origin of human music to humans imitating birds.<sup>81</sup> But using the word *music*, or the word *song*, about species other than our own is a tricky business. On the one hand, we might call sounds of some passerine species *birdsong* simply because it is produced with the animal's voice and is pleasing to us, and that naming decision might lead us to make unwarranted assumptions about what their vocalisation is to the birds that produce it. On the other hand, the concept of music can be, as the zoomusicologist Dario Martinelli laments, 'totalized by species, as strictly human-related,'<sup>82</sup> and then we lose a distinctive opportunity to understand where our own species fits in the broad context of animal life.

It is almost unavoidable that when we examine and describe features of other species that bear some resemblance to human features – whether anatomical, behavioral, or even cognitive – we do so by borrowing terms that we use about ourselves. We are then faced with the challenge of reminding ourselves that the term *song*, when applied to a sparrow, is a metaphor. Zoologists go further in applying human terms: they use pairs or sets of human metaphors to distinguish different features of another species. So it is when the vocalisations of 'songbirds' are sorted into *songs* and *calls*, or when the sounds that whales make are sorted into *songs* and *calls*, or *clicks* and *whistles*.<sup>83</sup> Zoologists evidently choose such sets of terms to distinguish kinds of sound patterns, as when the ornithologist William Thorpe in 1961 defined *bird-song* as 'a series of notes, generally of more than one type, uttered in succession and so related as to form a recognizable sequence or pattern in time',<sup>84</sup> thereby distinguishing those series from the shorter, simpler *calls*.

It is undoubtedly handy to have different terms for categorizing types of utterance that sound obviously different to human observers, utterances that may also serve different roles in the lives of the animals, just as it is handy for English-speaking humans to categorize their own utterances as *song* or *speech*. But what is handy is also apt to be misleading. The term *song* brings to our minds a rich human experience of sounds, activities, situations, cognitive processes, and states of mind – a substantially different experience from what the term *speech* brings to mind, and we also have experience of song and speech combined in a single stream. When we distinguish 'songs' from 'calls' in birds or whales, however, we cannot assume that our terms are marking two

<sup>81</sup> Lucretius, *The Nature of Things [De rerum natura]*, Book 5, ll. 1379ff., trans. A. E. Stallings (Penguin Books, 2007), 192.

<sup>82</sup> Dario Martinelli, *Of Birds, Whales, and Other Musicians: An Introduction to Zoomusicology* (University of Scranton Press, 2009), 217.

<sup>83</sup> See Laela Sayigh and Vincent Janik, 'Cetacean Communication', in *Deep Thinkers: Inside the Minds of Whales, Dolphins, and Porpoises*, ed. Janet Mann (University of Chicago Press, 2017), 72–76.

<sup>84</sup> W. H. Thorpe, *Bird-Song: The Biology of Vocal Communication and Expression in Birds* (Cambridge University Press, 1961), 15.

comparable ranges of experience for the animals in question. We can be confident that the 'calls' of these species, whatever their functions, cannot generate thought in anything like the way human speech does. It may also be that the more we learn about bird or whale 'song', the more unlike human song or speech it will appear in some or all respects. But we may still be able to learn, both about these remote species and about our own, from making the comparison.

One path to learning about other species may come from what the present study shows about the cognitive faculty for music in humans: that we form our music in ways conducive to the operation of that particular cognitive faculty. That could mean that if we found another species that produced sound patterns resembling our music in the relevant respects, we might posit that animals of that species are producing and listening to those sounds by means of a cognitive faculty likewise corresponding to our faculty for music. This possibility brings us back to the humpback whale, because recordings of its 'songs', besides beguiling generations of human listeners, have lent themselves with remarkable ease to processes of formal analysis long practised by humans on their own music, and yet gaining access to the cognitive processes of humpback whales is such a forbidding task that any clue to it seems worth pursuing. In undertaking this project we should be on guard against drawing false analogies between human and humpback processing of sounds and at the same time be prepared to identify underlying similarities that may be hidden by the vast differences in the two species' anatomies, behaviours, and conditions of life.

The complex utterances of humpback whales have been subjected to musical analysis (i.e., analysis of relativities and relationships in the sequence and flow of sounds) since 1971, when Roger Payne and Scott McVay published their landmark article 'Songs of Humpback Whales'.<sup>85</sup> There they describe how the males of a humpback population sing a shared 'song' during the months of each year when they are on their breeding grounds, songs up to half an hour in length that they then repeat 'with considerable precision' in 'song sessions' that may last for several hours. According to later research, a song changes considerably over the course of that season through a process of individual innovation and group adoption. But during the rest of the year, when they are on their feeding grounds and, as Katharine Payne reports, 'when there is very little singing, the song hardly changes: early songs on the breeding grounds are similar to those last heard in the previous season'.<sup>86</sup> Humpback singing therefore requires prodigies of memory across various lengths of time.

The whales form their songs on principles of segmentation and repetition that may aid their memory, as the same principles do for human musicians. Payne and McVay describe the humpback songs as formed in a hierarchy of segments, which they designate with musical terms: subunits, units, subphrases, phrases, themes, songs, and song sessions. At each level of this hierarchy, they find considerable repetition of

---

<sup>85</sup> Roger Payne and Scott McVay, 'Songs of Humpback Whales', *Science*, 173/3997 (13 August 1971), 585–97.

<sup>86</sup> Katharine Payne, 'The Progressively Changing Songs of Humpback Whales: A Window on the Creative Process in a Wild Animal', in *The Origins of Music*, ed. Wallin and others, 139.

the constituent segments, so that – as with much human song – a single song or song session is a web of repetitions. And the segmentation is made clear not only by those repetitions, but also by short pauses between segments at all levels. Subsequent study of these songs, collected over half a century from humpback groups spread across the world's oceans, has generally confirmed the usefulness of the analytic system created by the Paynes and McVay, requiring relatively little adjustment, because as Danielle Cholewiak, Renata Santoro Sousa-Lima, and Salvatore Cerchio reported in 2013, 'the overall hierarchical structure is observed globally, thus a heritable species-level characteristic, although the details of song patterns differ between populations of males that are acoustically isolated during all seasons'.<sup>87</sup> Nevertheless Eduardo Mercado and Stephen Handel had challenged that analytic system on several grounds in 2012, maintaining that determining the boundaries and configuration of segments is not straightforward (a problem that analysts of human music wrestle with too!) and that certain physical constraints on the whales' movements and air circulation produce 'acoustic regularities in the rhythm, timing, or spectral properties of sound production that may occur independently of the types of units being produced'<sup>88</sup> – that is, regularities that cut across the segments that Payne and McVay first reported. But nothing about this criticism disturbs the idea that the songs are formed from sequences of segments embodying repetition at various levels, a system that could be valuable for the whales' memory.

In fact, Mercado and Handel's attention to 'acoustic regularities' in the songs points us to something else that could bear on the whales' exercise of memory: the rhythmic and pitch stability in repeated units. Though the individual sounds in these songs are less sustained in pitch than the notes of much human singing, the frequency path of each sound tends to be very nearly the same from one iteration to another, and though the whale performers, like human performers, may vary the tempo of a song on different occasions, rhythmic relations within phrases or themes may be quite stable. Do these whale musicians measure time with a temporal grid or measure frequency with a pitch grid? It is hard to say, but it is certainly possible to connect their ability to replicate pitch and time structures from one repetition of a phrase, theme, or song to the next with an ability to overlay music from memory on music that 'is coming to be'.

One structural feature of humpback songs – the pauses between segments – deserves special consideration because those pauses can seem so 'human' to us and yet they point to how alien whales are to us in their anatomy, their way of life, the very medium they live in. The pauses seem 'human' because they remind us of the breaths that are so strongly connected to the segmentation of human song at the level of the phrase. But the pauses in humpback songs are not signs of breathing; they are frequent and incessant over the course of a single dive (though breathing tends to occur 'at a

---

<sup>87</sup> Danielle Cholewiak and others, 'Humpback Whale Song Hierarchical Structure: Historical Context and Discussion of Current Classification Issues', *Marine Mammal Science*, 29.3 (July 2013), E312–E332, E314.

<sup>88</sup> Eduardo Mercado III and Stephen Handel, 'Understanding the Structure of Humpback Whale Songs', *Journal of the Acoustical Society of America*, 132 (2012), 2947–50.

predictable point in the song').<sup>89</sup> So, how do the whales produce their sounds underwater? Recent discoveries about the mechanism of song production in humpbacks suggest that whales of that species vocalize by means of a larynx that 'functions by the same myoelastic-aerodynamic principles as the larynx in humans', but using air that circulates continuously between the larynx and lungs in the trachea, which is closed off from the outside atmosphere while the whale is underwater.<sup>90</sup> The process is therefore something like the circular breathing that some human woodwind players learn in order to play long passages without pausing for breath. The frequent pauses in the humpbacks' singing may be related to an alternation of vocalizing mechanisms as its air flow reverses direction,<sup>91</sup> or to the singers' process of recalling these long, complex strands of song, or to the need to help listening whales apprehend and respond to the songs across long distances in the ocean.<sup>92</sup> All that can be said for sure is that the pauses in the humpbacks' songs are definitely musical features, in that they are strictly coordinated with repeating segments of the songs.

The performance practice of humpback performers suggests something remarkable about the cognitive faculty they are employing. The males of a given population, gathered in a certain proximity to each other, give overlapping performances of the same song. As Katharine Payne writes, they repeat 'the same phrases and themes in the same order, but not in synchrony with one another'. What is more, 'in each population the songs were continuously and rapidly changing'.<sup>93</sup> That is, each singing whale listens to the others and picks up alterations here and there to the shared song, until the whole population is singing an altered shared version that itself keeps changing for the duration of that season. This process of 'cultural evolution' of a song, as Payne names it,<sup>94</sup> raises many questions about how the humpbacks' cognitive process works. Since they do not sing 'in synchrony with one another', they are not discovering divergences between the song they are singing and the altered version sung by a fellow singer from hearing those two versions sounding against each other. But when they are listening to a fellow singer, they must be comparing what that other whale is singing to what they are running from memory – and perhaps even when they are at a different point in performing the song themselves. That would be quite a feat for a human singer, but it is

<sup>89</sup> Katharine Payne and Roger Payne, 'Large Scale Changes over 19 Years in Songs of Humpback Whales in Bermuda', *Zeitschrift für Tierpsychologie* (now *Ethology*), 68 (1985), 91.

<sup>90</sup> C. P. H. Elemans and others, 'Evolutionary Novelty Underlies Sound Production in Baleen Whales', *Nature*, 21 (2024), Discussion. See also Olivier Adam and others, 'New Acoustic Model for Humpback Whale Sound Production', *Applied Acoustics*, 74 (2013), 1182–90.

<sup>91</sup> Elemans and others, 'Evolutionary Novelty'.

<sup>92</sup> Personal communication, Hansen Johnson, Dalhousie University, 14 March 2016.

<sup>93</sup> Payne, 'The Progressively Changing Songs', 138. These group alterations were first described in Katharine Payne and others, 'Progressive Changes in the Songs of Humpback Whales (*Megaptera novaeangliae*): A Detailed Analysis of Two Seasons in Hawaii', in *Communication and Behavior of Whales*, ed. Roger Payne (Westview Press, 1983), 9–57.

<sup>94</sup> Further research has shown that this 'evolution', in which the content of a song grows steadily more complex, is supplanted at times by 'revolutionary' events, in which songs are 'always completely replaced with a simpler song'. Jenny Allen and others, 'Cultural Revolutions Reduce Complexity in the Songs of Humpback Whales', *Proceedings of the Royal Society B*, 285 (2018), 3.

evidently just what the humpback faculty for music equips humpback males to do as a matter of course.

Study of the patterns of humpback songs has yielded some insights into the cognitive processes that could have produced those patterns. Linda Guinee and Katharine Payne, for instance, studied what they called ‘rhyme-like repetitions’ in the songs and noticed that those repetitions occurred most often when the songs were most complex, which is to say when the singers had the most material to remember. The two researchers, thinking like musicians as well as like biologists about what it would take to perform a long, complex song, proposed that these musical rhymes served a mnemonic purpose.<sup>95</sup> In that observation they reminded us that learning and remembering and revising and reproducing music, especially on such a scale of time and complexity and with the inclusion of recent alterations, takes tremendous powers of sustained concentration – concentration on the music itself. That thought prompts us to ask: if humpback males need to concentrate for long stretches of time on the intricacies of what they are singing, does their absorption in that task lead them to lose their consciousness of the here and now to a greater or lesser degree, as concentrating on music-making can do for humans? And if so, how can they give their attention at the same time to whatever here-and-now social functions the singing serves for them? It is difficult to observe the social interactions of humpback whales, and researchers have debated how the songs may figure in sexual display, competition among males, or group solidarity.<sup>96</sup> One kind of social interaction is indisputable: we know the songs serve for social exchange among the singing whales of a group, because they are listening to and adapting to each other’s performances. In all, the relationship between individuals’ use of their faculty for music and the functions to which they put their music appears to be as rich and complex a field for continuing inquiry in the case of humpback whales as it is in the case of human beings.

Aristoxenus describes the faculty for [comprehending] music as if it were a routine process: ‘we have to perceive what is coming to be and remember what has come to be’. But that faculty may seem routine to us only because we are humans and it is an element of our cognitive makeup. Perhaps in the whole animal kingdom it is a somewhat rare trait.<sup>97</sup> Other species, including some such as chimpanzees and bonobos

<sup>95</sup> Linda Guinee and Katharine Payne, ‘Rhyme-like Repetitions in Songs of Humpback Whales’, *Ethology*, 79 (1988), 295–306. See also Payne and others, ‘Progressive Changes in the Songs’.

<sup>96</sup> Louis Herman, ‘The Multiple Functions of Male Song within the Humpback Whale (*Megaptera novaeangliae*) Mating System: Review, Evaluation, and Synthesis’, *Biological Reviews of the Cambridge Philosophical Society*, 92 (2017), 1795–818. The situation is described succinctly – ‘While song is well described, its function is not yet fully understood’ – by Dana Cusano and others in ‘Socially Complex Breeding Interactions in Humpback Whales Are Mediated Using a Complex Acoustic Repertoire’, *Frontiers in Marine Science*, <https://doi.org/10.3389/fmars.2021.665186>.

<sup>97</sup> A kind of comparison across time is found even in bacteria, in which, according to Peter Godfrey-Smith, ‘one mechanism registers what conditions are like right now, and another records how things were a few moments ago’. Peter Godfrey-Smith, *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness* (Farrar, Straus and Giroux, 2016), 17. But that capacity does not involve anything like the comparison of simultaneously remembered and perceived sound sequences required in musical processing.

that are closely connected to us in evolutionary terms and undoubtedly clever in many ways that we recognize as clever, perhaps do not possess anything like our capacity to make comparisons of sounding sequences across time. Finding that certain species' utterances, including humpback songs, have structural features that seem adapted to a faculty for music at all like ours could give researchers a clear criterion for determining which species can reasonably be studied as 'musical'. The faculty itself could then be studied not as a uniquely human trait (which the faculty for language is generally taken to be), but as a faculty shared across a scattering of species, which intriguingly may well include some from different classes of animal, as well as some that dwell on land, some in the water, and some in trees and air.

The model of the human faculty for music proposed here is a propensity to compare memory to perception in sounding sequences by creating measured order in time and pitch and by creating pattern through segmentation and repetition. In human musical experience, we can find some consequences of that propensity in our responsiveness to the kinds of sonic order and patterning we consider beautiful, in our ability to become obsessed by the realization of those qualities, in our ability to excite or soothe our own spirits and share spirits with others by absorbing those qualities into our bodily motions and responses. In short, the study of the human faculty for music and its matching to human music has the potential to lead to insights about human cognition, consciousness, behaviour, culture, and evolution that the study of music alone cannot provide. To the extent that we can turn that study into a cross-species investigation, we can use what we are learning about ourselves to probe the cognitive processes of at least a few other species, even as we can come to consider our own artistic creativity – at least the musical side of it – a 'natural' result of cognitive abilities we share with those species.