

The Evolution of Gender Dimorphism in the Human Voice

The Role of Octave Equivalence

Nicholas Bannan, Robin Dunbar, and Joshua Bamford

Humans exhibit what appears to be a unique vocal property: octave equivalence, whereby adult male voices are, on average, an octave lower in pitch than those of adult females and children. The evolutionary significance of this seems largely to have escaped notice. While sexual selection might explain why male voices are generally lower, it cannot explain why they should be so much lower than what would be expected for body size or why the average difference should be exactly one octave. Nor does a generalized dimorphism convey why precisely tuned octaves feature so commonly in human vocal interaction. The octave features strongly in the organization of music. A consequence of this characteristic of human pitch perception and production is the capacity to share and respond to vocal pitches (and their instrumental equivalents) as if they are “the same” irrespective of the difference in range, a phenomenon known as octave equivalence. We investigate the nature of octave equivalence from an adaptive perspective and propose a hypothesis for its evolution based on the importance of chorusing for social bonding and pitch matching in intergenerational exchange.

Online enhancements: appendixes.

Introduction: Prevailing Themes in the Evolution of Musical Communication

An increasing body of research in a variety of disciplines (Bannan 2003, 2008, 2020; Dunbar 2012*a*, 2014; Harvey 2017; Morley 2013; Savage et al. 2021; Wallin 1991) has converged on a position that presents the human capacity for music as adaptive and as representing the potential bridge between animal communication and human language. Features of protomusical behavior that contribute to this perspective include the capacity, unmatched in any other species, to entrain to a beat and to achieve this during both locomotion and work with the hands, as well as in dance (Larsson, Richter, and Ravignani 2019); the unique capacity to match pitch precisely as a potentially meaningful interaction with others (Podlipniak 2016); the similar capacity to match timbre and innately to perceive connections between the properties of timbre and frequency (Sethares 1993; Titze 2003); the capacity to share common purpose in utilizing these abilities to communal benefit (Dunbar 1998) and to promote social bonding (Dunbar 2012*a*; Savage et al. 2021; Tarr, Launay, and Dunbar 2016*a*); the capacity to utilize these abilities to act on memory and set up expectations for subsequent events through mediating the experience of the past in the present and shaping the response to the future (Druskin 1983:120; Podlipniak 2020; Stravinsky

1962); and the capacity to affect, elicit, and convey emotion (Belyk and Brown 2016; Scherer 1992; Snowdon, Zimmermann, and Altenmüller 2015). Few of these behaviors are apparent in our closest primate relatives to the extent that they are in humans (Harvey 2017; Merker 2000, 2012) and some not at all.

Within this broad pattern, however, one observation stands out: in humans, the two sexes differ in their voice pitch, or fundamental frequency, both when speaking and when singing. Comparative cross-species studies of vocalization frequency have noted a consistent negative relationship between fundamental frequency and body size. Larger species have both a lower fundamental frequency (F_0) and lower formants ($F\Delta$), on average, than smaller species, a relationship that is particularly strong among the primates (Bowling et al. 2017*a*). In many mammals, the males have both a lower F_0 and $F\Delta$ than females (Charlton and Reby 2016; Fitch and Hauser 2003). This sexual dimorphism is more pronounced in highly territorial species and those with polygynous mating strategies (Charlton and Reby 2016; Puts et al. 2016). Studies of individual differences in humans suggest that certain traits (such as sex, size, and upper-body strength) may be predicted by the frequency profile of one’s vocalizations (Pisanski, Groyecka-Bernard, and Sorokowski 2021; Pisanski, Raine, and Reby 2020; Pisanski et al. 2014). Crucially, listeners can infer these

Nicholas Bannan is Honorary Research Fellow at the Conservatorium of Music of the University of Western Australia (35 Stirling Highway, 6009 Perth, Australia [nicholas.bannan@uwa.edu.au]). **Robin Dunbar** is Emeritus Professor in the Department of Evolutionary Psychology of the University of Oxford. **Joshua Bamford** is a DPhil student at the Institute of Cognitive and Evolutionary Anthropology of Oxford University. This paper was submitted 6 IX 20, accepted 2 XII 21, and electronically published 7 VI 24.

traits from hearing another's voice (Raine et al. 2019); they can, for instance, even predict the outcome of a tennis match by comparing the F_0 of grunts made by the players (Raine, Pisanski, and Reby 2017). Low-frequency vocalizations seem to predict dominance in human males rather than attractiveness to females, suggesting that the function of vocal sexual dimorphism in humans may be for male-male competition (Puts et al. 2016). This has inevitably resulted in a widespread assumption that vocal dimorphism is simply a consequence of mating competition and, hence, sexual selection.

While sexual dimorphism is most often associated with polygynous mating systems in primates, the sexual dimorphism for F_0 in humans is far greater than in most other primates (Puts et al. 2016) and is the greatest of any ape (fig. 1), despite humans not being strictly polygynous (Schacht and Kramer 2019)—and in contrast to humans' consistently intermediate position between monogamous and polygamous primates on most anatomical indexes of mating system (including body size dimorphism, relative testis size, and 2D4D ratio; Harcourt et al. 1981; Plavcan and van Schaik 1997; Wlodarski, Manning, and Dunbar 2015). More puzzling still is the fact that the voice pitches of the two sexes are, on average, *exactly* an octave apart in humans, especially when singing. More importantly, humans appear uniquely to possess and exploit a phenomenon in pitch perception known as octave equivalence (OE; Hoeschele, Weisman, and Sturdy 2012)—the ability to recognize that two notes are in harmony when an octave apart. Not only does this seem to have been little noticed by evolutionary scientists (though it is well known to musicologists) but also there appears to be no obvious explanation for it. While we can accept that a lower male pitch is the result of sexual selection, as it is in other

animals, some additional selection pressure is required to explain why the voice of the human male permits vocalization exactly an octave below that of the female.

This effectively creates a three-stage explanatory structure (the evolution of musicality, the evolution of a sex difference in pitch, and the evolution of OE) that may each involve different historical selection pressures. Our concern here is with the third. We concur with the suggestion that musicality (and singing, in particular) evolved as a generic community bonding mechanism (Dunbar 2012a, 2014, 2016; Dunbar et al. 2012; Savage et al. 2021) and likewise with the suggestion that the existence of a sex difference in pitch is due to sexual selection (probably as a correlated trait of divergent body sizes; Apicella, Feinberg, and Marlowe 2007; Darwin 1871; Dunn et al. 2015; G. Miller 2000a; Puts, Gaulin, and Verdolini 2006). What remains to be explained is why the sexual dimorphism in pitch is exactly one octave. If sexual selection is the key driver for hyperlowered pitch in males, why did male-male competition for access to females not drive it even deeper? Indeed, why has the sexual dimorphism in pitch not reduced, since humans do not primarily have an obligate polygynous mating style? The fact that the difference is exactly one octave is too precise to be an accident and instead implies that something else acted as a break on runaway selection. We will argue that the answer is group bonding during communal chorusing (with dyadic bonding as a secondary possibility). To make this claim, we first provide an overview of OE and its anatomical basis and then make a case in the light of this for social bonding as the likely function. Finally, we ask whether this evolved early or late in hominin evolution and offer some indirect evidence for a likely date. We should be clear that, in terms of Tinbergen's (1963) "Four [biological] Whys," our main concern is with the why and when, rather than the mechanisms and ontogenetic "hows," though we will comment on both of the latter in passing. We conclude by discussing how future research might illuminate and test this suggestion.

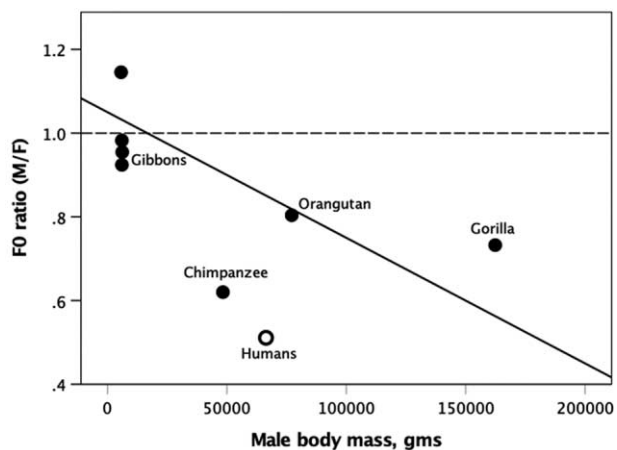


Figure 1. Relationship between body mass and vocalization pitch in the higher primates. The dashed line indicates male fundamental frequency (F_0) = female F_0 ; the solid line indicates where the ratio would lie if it directly reflected sex dimorphism in body weight. Polygamous species have $F_{0\text{ male}} < F_{0\text{ female}}$. Within the great apes (nongibbons), human males have much lower F_0 than females, notwithstanding their smaller body size compared to many other great apes. Adapted from Puts et al. (2016).

Octave Equivalence: Ontogeny and Phylogeny

Prepubescent children of both genders have almost identical vocal anatomy from before birth until the age of 10 or so (Fitch and Giedd 1999; Sergeant, Sjölander, and Welch 2005; Weiss 1950). The vocalization of these prepubescent children is achieved with close to identical anatomical configurations. Indeed, expert listeners are often unable to discriminate with certainty between the singing of individual boys and girls (Sergeant, Sjölander, and Welch 2005) or between boys', girls', and mixed choirs (Welch and Howard 2002). Both sexes undergo a similar primary descent of the larynx within the first year of life, as the need simultaneously to breathe and to feed evident in initial suckling gives way to growth within the vocal tract (Fitch 2002). However, there is a marked divergence early in adolescence.

From adolescence onward, however, maturing males diverge progressively from male children, as well as all females,

eventually achieving a pitch as adults that is approximately an octave lower—an expression of the phenomenon referred to as octave equivalence (see fig. 2). These developmental changes are universal across human societies (Titze 2003, 1994), consequent on hormonal and anatomical causes that are determined by our genes, though there are slight differences in the age of onset of puberty as a result of environmental factors (de Muinck Keizer-Schrama and Mul 2001).

Concurrent with the acquisition of vocal fluency in speech and song is the development of registers (Henrich 2006), acoustic phenomena determined by varying discontinuities in the ranges within which the relationship between fundamental frequency and optimum resonance are available (Cooksey 1997; Gackle 1991; Story, Titze, and Hoffman 2001). While experienced singers can learn both to overcome “breaks” between registers (D. Miller 2000) and to exploit them in vocal styles such as yodeling (Echternach, Markl, and Richter 2011), most adults remain constrained by their natural (modal) vocal register. The principal developmental trait that requires explanation is the approximate octave difference between the modal voices of adult men, on the one hand, and, on the other, women and children (Aristotle 1927; Hoeschele 2017; Peter, Stoel-Gammon, and Kim 2008).

Importantly, vocal anatomy does not suggest that this developmental process occurs for purely physical reasons (Fitch and Giedd 1999; Titze 2003). The crucial factor in achieving this unique result resides in the neural control of production and perception, without which it would fail (Harvey 2017; Titze

2017). Form may well have preceded function, but the latter relies upon brain development that enabled the otherwise unpromising engineering of the vocal tract (Titze 2003) to achieve song, language, and culture. Consequently, OE in humans embraces three related phenomena: the perception of OE; the ability to produce octaves vocally, in imitation as well as simultaneous performance; and the average octave difference between male and female voices. In sum, compared to species such as gibbons with similar vocal production across genders that permits equal duetting and turn taking (Pika et al. 2018), a clear dimorphism evolved in humans. It is reasonable to assume that this represents a development whereby our species saw its males undergo change by stages, whereas female voices remained more or less as they were in childhood. At a point in this transformation, a new and beneficial factor was encountered. As male voices reached a range an octave lower than females, they became capable of interacting in octave unison (Bannan 2020), allowing them to leverage the perception of OE. At this point, we suggest, male voices became tethered to female voices—not in the same range but in an equivalent range that permitted imitation and reinforcement that exploited the harmony made possible by the octave difference.

OE has been detected as an aspect of auditory perception in species other than humans (e.g., rhesus monkeys [Wright et al. 2000], white rats [Blackwell and Schlosberg 1943]), though only rarely in animal sound production. A dolphin learning to imitate synthesized pitch contours was able to achieve this through transposing its vocal production to a range it was able

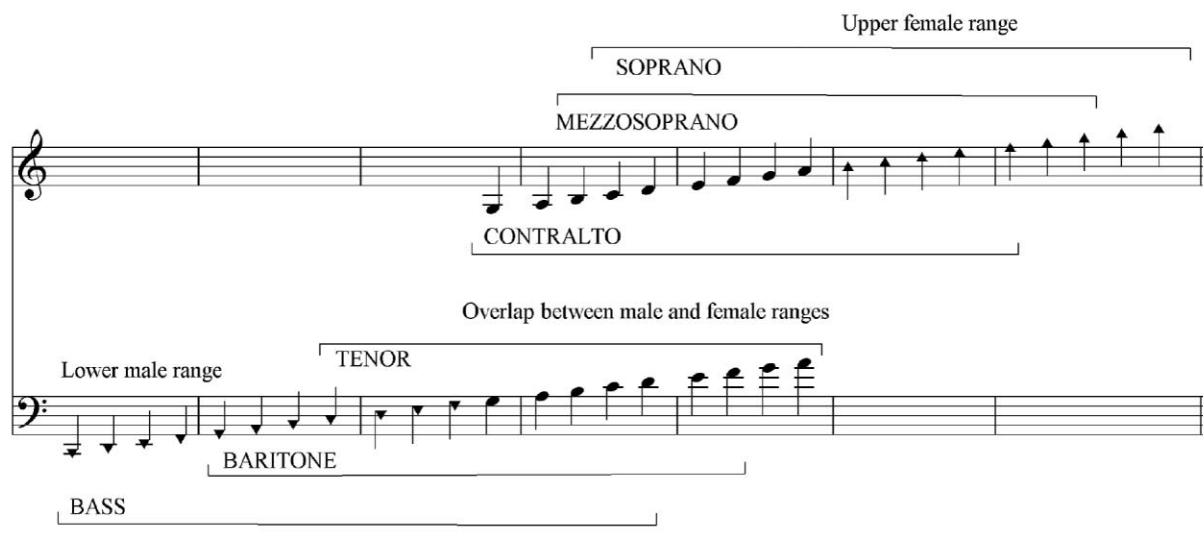


Figure 2. Variety of ranges within human singing. Children’s voices map, on average, onto the mezzo-soprano range. Note both the phenomenon of octave equivalence and the overlap between the upper range of male voices and the lower range of females and children. The onset of puberty introduces a phase in which maturation in girls is associated with relatively minimal change (Gackle 1991; Weiss 1950), while that of boys determines a period of descent and growth that, over time, leads to a drop in fundamental frequency of about an octave and a related change in formant distribution (Cooksey 1997; Fitch and Giedd 1999; Harries et al. 1997). This is associated with a second laryngeal descent (Fitch 2002; Sergeant and Welch 2009; Weiss 1950) and enlargement of the vocal folds that commences in the midstage of puberty (Cooksey 1997), eventually conferring a voice sounding approximately an octave lower than its prepubescent form. It is worth noting in this context that young boys appear to adopt, quite instinctively and unconsciously, lowered formants when singing in the presence of girls (Keller, König, and Novembre 2017).

to access (Richards, Wolz, and Herman 1984). Attempts to elicit OE in avian vocal mimics such as budgerigars (Wagner et al. 2019) and chickadees (Hoeschele et al. 2013) have, however, failed, suggesting that it may be an exclusively mammalian characteristic. Development of behaviors dependent on the perception of OE would thus appear to be a special human trait (Hoeschele, Weisman, and Sturdy 2012) where perception and production represent a reliable feedback loop (Hockett 1960; Jones and Keough 2008). Doğantan-Dack (2013:212) concluded that “the emergence of tonal encoding of pitch can be construed as a pre-linguistic stage . . . intimately related to the evolution of affective capacities.” We provide further material concerning gender, vocal development, and performance in appendix S1 (apps. S1–S6 are available online).

Theoretical investigation of the phenomenon of OE and derivation of the frequency ratios of the harmonic series (fig. 3) have been known since Pythagoras (Hagel 2009) and, independently, the ancient Chinese (Falkenhausen 1992). OE in the implicit practices of musical performance has been found all over the world (Hoeschele, Weisman, and Sturdy 2012) and is viewed as salient in musical cognition (Borra et al. 2013; Deutsch and Boulanger 1984; Savage et al. 2021). Neural response to pitched sounds suggests that harmonic series relationships have shaped human hearing and aural engagement (Bowling, Purves, and Gill 2018; Foss, Altschuler, and James 2007; Warren et al. 2003). Aristotle remarked (*Prob. Xix 39*) on the perceptual and developmental properties of OE, and the phenomenon has underpinned the construction of musical instruments and compositional techniques throughout modern history (Berlioz 1844; Mersenne 1636; Praetorius 1958 [1618]; Rameau 1722), predominantly dependent on a 2:1 ratio in size.

Acoustic analysis of large corpora of vocal utterance (Chiba et al. 2019; Kuroyanagi et al. 2019; Schwartz, Howe, and Purves 2003) confirms that human vocalization, especially singing, maps with considerable accuracy onto harmonic series properties and the contours of melodic intervals derived from them

(Robledo et al. 2016). A single example of the phenomenon must suffice to illustrate this. The pentatonic scale (Do, Re, Me, So, La—in solfège notation) has been claimed as a human universal (Trần Van 1977). In Tuvan overtone singing, it can be heard as deriving from (or mapping onto) harmonics 8, 9, 10, 12, and 13 above a sustained fundamental, or drone (Trần Quang 2002; see fig. 3). Melodies performed over a drone are also a near universal (Gabisonia 2015; Jordania 2006; Nikolsky 2015; van der Merwe 1989; see also Parncutt 2012), according to Podlipniak’s (2016) definition of “pitch centrality,” which implies it is the oldest and most robust form of harmonic accompaniment. This suggests that human singing is well characterized by the use of harmonics and has been so for a long time.

A potential role for OE in language acquisition, in which children imitate the voices of both female and male ranges (Peter, Larkin and Stoel-Gammon 2009; Peter, Stoel-Gammon, and Kim 2008), implies a more general, innate response than has hitherto been recognized. In parallel with this, traditional exploitation of harmonic series properties (i.e., not explicitly informed by theoretical understanding) has been found in sophisticated and varied forms in a wide variety of musical practices (Bannan 2008; Dargie 2018; Ikhtisamov 1988; Trần Quang and Bannan 2012). Perceptual OE that may represent an adaptive response to the properties of the harmonic series would seem to have shaped human productive capacity to accurately map vocalizations onto those an octave higher or lower. We summarize additional evidence for the phenomenon of OE in human vocalization in appendixes S2–S4.

Conventional Explanations for the Evolution of Human Vocal Development

Body size alone would explain a lowered voice in males and the consequent vocal dimorphism between the two sexes. This would likely reflect the outcome of sexual selection driven by

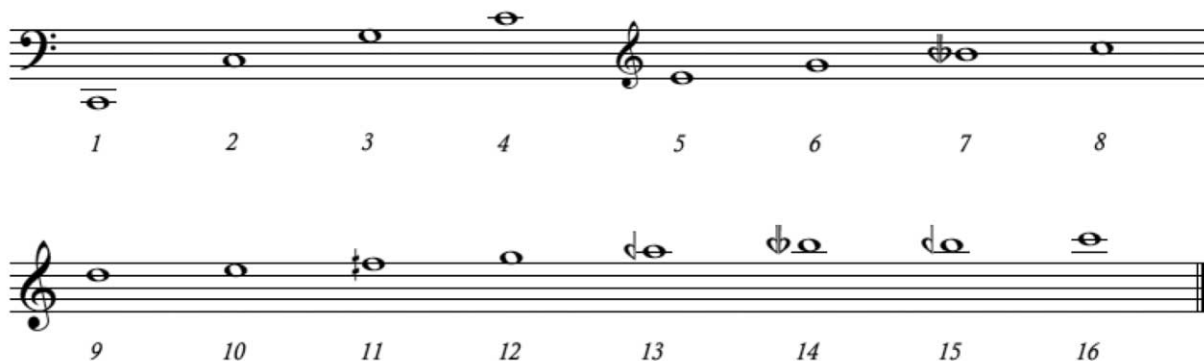


Figure 3. Natural phenomenon of the harmonic series set out sequentially in music notation. The numbered pitches indicate the frequency ratios by which any two harmonics can be compared so as to derive the musical interval between them. Thus, 2:1 is the octave whose equivalence is central to the argument of this paper (note that 2:4, 3:6, etc., also result in octaves, while 3:4 results in a perfect fourth, and so on). Formants derived from harmonics sounding in relation to fundamental frequency (F_0) are perceived as distinct vowels. In overtone singing, the fundamental is sustained while oral filtering amplifies individual harmonics to produce melody.

male-male competition and/or male attractiveness to females (Apicella, Feinberg, and Marlowe 2007; Dunn et al. 2015; G. Miller 2000a; Puts, Gaulin, and Verdolini 2006). Low pitch in a variety of species has such functions (Charlton and Reby 2016; Fischer et al. 2002; Fitch and Reby 2001), and this may be selected for partly through the iterative preference of females for low voices as evidence of good providers with desirable genes (Saxton et al. 2016). This may well explain a developmental difference, but why in humans does the difference take the form of a precise octave that allows harmonic equivalence? Why not a runaway process (Zahavi and Zahavi 1999) resulting in a difference in range of *more* than an octave? Or, why, for that matter, would a smaller interval of difference not have been sufficient?

Cross-species comparisons (Ghazanfar and Rendall 2008; Honing et al. 2015) identify the allometric constraints for range and vocal character that illuminate the anatomical and acoustic features that characterize modern human vocality and its pathology (Sundberg 1987; Titze 1994). Even allowing for possible differences in size between males and females as a factor in human survival that was selected for by nonsexual drivers such as defense or hunting, the average bodily size difference of ca. 10%–20% (Ghazanfar and Rendall 2008) does not correlate with the 2:1 ratio of the difference in average fundamental pitch (F_0) and the corresponding average size difference in vocal fold length (2 cm in males vs. 1 cm in females; Titze 2003). For it to do so, human males would need to be some 3 m tall.

An alternative possibility is the use of song in male-male bonding—probably a near-universal cross-cultural phenomenon. Such a function for chorusing was proposed for archaic human males by Merker (2000), and the harmonic potential of mutual reinforcement that universally characterizes human vocal interaction has been identified and explored by Bowling and Purves (2015). However, patterns of exclusively male/male cooperation in musical vocalization (Faulkner 2012) clearly differ from those associated with female/female relationships (Garnett 1999) and the activities exhibited as anthropologically proper to each (Dávid-Barrett and Dunbar 2013; Gaulin and Boster 1985; Knight and Lewis 2017). While male collective and competitive chorusing on the lines proposed by Merker (2000) surely represents a potential stage in the development and function of human vocalization, the instrument required to perform it—the deep male voice—had its anatomical origins in one originally pitched higher, within the female range, from which it departed under sexual selection. However, this does not explain why male voices have the pitch they do: if male bonding were the sole issue, any pitch would suffice. More importantly, perhaps, focusing solely on male voices simply leaves us bereft of any reason why women and children should have voices too—let alone why they are clearly different. It is therefore essential to account for the contributory role of the voices of females and children in any model of vocal evolution.

An alternative possibility that might circumvent this difficulty is that vocal exchanges might underpin pair bonding.

Outside the obligately monogamous gibbons, humans are unusual among the apes in exhibiting a strong (albeit usually temporary) form of pair bonding (i.e., facultative, as opposed to obligate, monogamy). Coordinated dancing, for example, is known to facilitate pair formation in at least some human communities (Pitcairn and Schleidt 1976). Hughes, Dispenza, and Gallup (2004, 2010) analyzed the relationship between opposite-sex vocal attractiveness ratings in adults, indexes of bodily dimorphism, and self-reporting of sexual behavior to propose that voice difference may have been an important parameter of mate choice in human evolutionary history. Weusthoff, Baucom, and Hahlweg (2013) interpreted the reproductive stage in the human life cycle not only in terms of the role of the voice in mate attraction in adults but also on the maintenance of paired relationships. They reviewed five available studies of the role of vocal F_0 in the vocally encoded emotional arousal of interactions between couples—what leads to friction in relationships and how this can be considered a component of relationship therapy. Their findings suggest that the means by which F_0 acts as an index of emotional arousal and of the intensity of emotional reaction are influenced by both basic biological processes and socially learned communication behaviors stemming from a similar evolutionary basis. At the same time, the achievement of precise tuning might, of course, represent (or be exploited to represent) an investment of stability, authenticity, and salience in vocal signals intended to convey credibility (Mehr et al. 2020), which could be underpinned by the achievement of precise tuning at the unison or octave. Such rich acoustic information carried by the varied timbral content of vocal signals involves a positive feedback loop by which theory of mind (Livingstone and Thompson 2009) would have developed and contributed to human social culture.

Plausible as this might seem, in fact most animal species where pairs call to each other to reinforce the pair bond (or maintain contact with each other) typically have vocal registers that are very similar (e.g., the African bell shrike, whose two-note duets are so tightly timed and individually indistinguishable as to sound like a single individual; Thorpe 1973). Moreover, in all these cases, the species are monomorphic, and the two sexes do not differ in vocal range, making this an unlikely explanation for humans. More importantly, if duetting did evolve to facilitate bonding, we would expect duet singing in the home rather than the communal singing that is so characteristic of human societies.

Finally, Falk (2004) argued that language evolved out of mother-infant singing and might thus have originated as a child-rearing strategy that allowed hairless females to free their hands for foraging and other work. Dissanayake (2008) built on this by suggesting that the origins of music as a parental response to infants rests on the acoustic and rhythmic features of mother-child bonding such that the gamelike aspects of adult sexual lovemaking would have been a natural extension of this for sustaining relationships. She describes a process whereby the mother-infant dyad displays bonding based on vocal and

physical processes that recapitulate in the adult female-male dyad:

The kinesic actions of mothers—touching, stroking, patting, holding the hand, embracing, grooming, and hugging their infant, and with an open mouth poutkissing against its body—echo a large number of the affiliative social behaviours noted in wild and captive primates . . . as well as being common adult gestures in humans of sympathy and affection. (Dissanayake 2008:175)

Identifying the husband-father role in the human family as unique among primates, she continues:

I suggest adding the hypothesis that in human courtship the evolved proto-musical capacities and sensitivities that enabled mother-infant interaction, especially as they promoted temporal coordination and contingent responsiveness, were exapted and transformed into the elaborated and prolonged loveplay that also distinguishes human sexual behavior from that of other primates. (Dissanayake 2008:185)

We consider some further implications of Dissanayake's model in appendix S5.

A Role for Octave Equivalence in Community Bonding?

Most of these claims focus on aspects of dyadic relationships. In common with many analyses of human behavior, they overlook one crucial feature of human and primate sociality, namely, the fact that all anthropoid primate species live in bonded social groups (Dunbar and Shultz 2021*b*). Such effects are underpinned by mutualism (Wrangham 1982) or group-level or group augmentation or group selection (Kingma et al. 2014; Kokko, Johnstone, and Clutton-Brock 2001; Okasha 2006), sometimes identified as a third component (social selection) of natural selection in addition to conventional survival-related and sexual selection (Crook 1972; Hamilton 1971; Lyon and Montgomerie 2012; Rubenstein 2012). For primates, the evolution of bonded social groups that are able to maintain spatial cohesion through time and space has been key to their evolutionary success, with the mechanisms that underpin group coordination evolving long before the evolution of the capacity to cooperate (Dunbar and Shultz 2021*b*; Shultz and Dunbar 2022). Group bonding is not an end in itself but a means to a fitness end (something that often seems to get overlooked in discussions of possible social functions; e.g., Mehr et al. 2020).

The central problem faced by all primates is the need to ensure that social groups remain coherent through time and space in the face of stresses that for all other mammals result in group fragmentation and dispersal (Dunbar 2020; Dunbar and Shultz 2021*b*). These stresses should not be underestimated; they have very significant effects on health and, in particular, on female fertility (Dunbar and Shultz 2021*a*). These latter effects are so strong that they make it impossible for mammals in general to live in stable groups larger than 15–20 individuals.

Anthropoid primates have been able to breach this glass ceiling by building bonded relationships by means of social grooming that buffer individuals against these stresses and counteract the centrifugal forces that otherwise act on them (Dunbar and Shultz 2021*a*). It is important to appreciate that bonding has nothing to do with cooperation (though bonding may give rise to cooperation in other contexts). Rather, its sole function is to maintain the integrity and cohesion of the group, so that the group can provide the protection against external threats (principally, predators and conspecific raiders) that would otherwise sink individuals' abilities to survive and reproduce in high-risk habitats.

Because there is a limit on the time that can be devoted to social grooming and other forms of affiliative behavior (Dunbar 2009) and because relationship quality is directly related to the time invested in it (Dunbar 1988; Sutcliffe et al. 2012), there is an upper limit on the size of primate groups that can be bonded in this way, at around 50 individuals (corresponding to the third layer of human social networks; Dunbar 2020, 2022). This time constraint is further limited by the fact that grooming is a strictly one-on-one interaction that, even in humans, cannot be extended to multiple recipients without a significant loss of intimacy and, hence, bonding effect. Since early hominin time budgets were already overstretched (Dunbar 2016), the only way group size could increase beyond this limit was to find behaviors that, in effect, allowed more individuals to be "groomed" simultaneously without need for physical contact (Dunbar 2012*a*, 2016). Vocal mechanisms (laughter and, later, chorusing or wordless singing) appear to have been the earliest of these to evolve (Dunbar 2012*a*, 2016, 2022). Both laughter and singing trigger the same neuropharmacological bonding mechanism (the brain's endorphin system) as social grooming and precipitate a sense of bonding to those with whom one does the activity (Manninen et al. 2017; Pearce, Launay, and Dunbar 2015; Pearce et al. 2015). Singing together seems to be uniquely capable of bonding individuals whose previous experience of each other has been entirely one of rivalry and competition (Pearce et al. 2015). In addition, singing appears to be uniquely scalable, allowing choirs of 200 or more to feel more bonded even than choirs of 20 (Weinstein et al. 2016). None of the other behaviors that we use in social bonding, including laughter, dance, feasting, and storytelling, seem to be capable of working on such a large scale.

We suggest that, in this context, there is something especially engaging about singing in unison, as well as in other well-tuned intervallic ratios derived from harmonic series relationships, and that doing so enhances the sense of bonding as a result (see also Savage et al. 2021). Of the most clearly perceived intervals derived from the harmonic series, it is the octave that has driven the most common and adaptively significant acoustic difference between the voices of men and women, evident in both song and speech. In a species capable of collective synchrony, precise attunement both efficiently amplifies the signal and preserves its integrity, permitting it to be retained in memory and shared accurately over time, thus enhancing bonding

between generations. Given that male voices had diverged away from female voices under the pressure of sexual selection acting on body size throughout pongid and early hominin evolution, there would, we suggest, have been a significant secondary selection pressure from social bonding to push male voices as fast as possible to the next nearest “sweet spot” at the lower octave.

Evidence for the Instinctive Achievement of Harmonic Interaction

What evidence do we have that harmonic singing of this kind enhances social bonding? Although this has not, as yet, been examined in any experimental studies, there is considerable circumstantial evidence from music practice that it is likely to do so and considerable experimental evidence that singing in general both influences social neurohormone titers and elevates the sense of belonging to a group. We summarize briefly some of this evidence here.

An intensified capacity for social bonding in complex a cappella harmony singing was reported by participants in Averill’s (1999) study of barbershop quartet performance. The same relationship between “just intonation” and style in historically informed performances of European Medieval and Renaissance vocal music has been identified by Covey-Crump (1992) in the rehearsal practice associated with solo voice consorts that specialize in this repertoire.¹ Precise tuning arouses heightened responses both in audiences and in the performers themselves. In the folk style of Sardinian male voice performance, Castellengo, Lortat-Jacob, and Léothaud (2001) report the phenomenon whereby four men singing together in precise just intonation can appear to elicit a “fifth voice.” In barbershop circles, this evidence of precise harmonic reinforcement is referred to as “the angels joining in” (Anonymous, personal communication, Barbershop Pioneer’s Convention, Elk Grove Village, Illinois, 1992) and is experienced as a peak aesthetic sensation both by the performers themselves and by listeners.

Singing in groups has been widely demonstrated to trigger the pleasure centers associated with endocrine expression (Harvey 2020), the release of endorphins (the principal bonding neuropeptide in primates; Pearce, Launay, and Dunbar 2015), the reduction of stress measurable in cortisol secretion (Fancourt, Aufegger, and Williamon 2015), and an enhanced sense of bonding (Pearce, Launay, and Dunbar 2015; Pearce et al. 2015; Weinstein et al. 2016). Moreover, we know from studies of other group activities (including rowing and dancing) that synchronized performance in groups raises endorphin upregulation significantly above that achieved by individuals performing alone (Cohen et al. 2010; Lewis and Sullivan 2018; Tarr, Launay, and Dunbar 2016b; Tarr et al. 2015). Such neural and limbic correlates of well-tuned singing cannot be mere side ef-

fects of capacities that have evolved for other purposes. On the contrary, we argue that they are adaptive traits that are foundational to the complex social and anatomical achievement of human group singing, the part this plays in collective social cohesion as well as interpersonal bonding—both mate selection and mate retention—and the transmission of vocalization between generations both in child-rearing and in fostering and upholding the identity of the wider family.

Dating the Origins of Human Singing

Our claim, then, is that because of the constraints imposed by time on the size of group that can be bonded by conventional social grooming, singing evolved to supplement grooming as a social bonding mechanism so as to enable humans to live in larger groups. If we could determine when a dramatic increase in grooming time became necessary, we might be able to estimate when in the course of hominin evolution the capacity to sing evolved. Since group size correlates closely with brain size (and especially neocortex size) in the hominids (Dunbar and Shultz 2021b), we can use cranial volume to determine group size, and then from this we can estimate the corresponding grooming time requirement (in each case using empirically determined equations; see Dunbar 2009, 2016, 2022). Figure 4 plots the estimated grooming time for each hominin species, along with equivalent estimates determined in the same way for the Old World monkey genus *Macaca* (macaques) and the ape genus *Pan* (chimpanzees). (Details for these calculations can be found in app. S6.) These data suggest two likely points at which grooming time demand underwent a significant increase: a smaller increase around 2.5 Ma with the appearance of early *Homo* and a much larger one around 500 ka with the appearance of archaic humans (leading to a further small increase with the appearance of anatomically modern humans, *Homo sapiens*, around 250 ka). To decide between the two main options, we need to look for anatomical signatures of the kinds of voice control needed for language and/or singing.

No primate devotes more than 20% of its day to social grooming (fig. 4, horizontal dashed line), and figure 4 suggests that none of the australopithecines would have exceeded this. Similarly, the pressure to evolve a novel bonding mechanism is at best modest in early *Homo* but reaches serious proportions for archaic humans (cf. the magnitude of the two dashed arrows in fig. 4). Singing (and speech) is dependent on a number of core features that include control over breathing (to allow for long exhalations without drawing breath), control over articulation (to vary pitch and timbre), and hearing (in order to match sounds between individuals). A number of relevant anatomical markers for speechlike vocal articulation are known to differ significantly between humans and other primates. Figure 4 plots, across the top of the graph, the known presence of nonhuman (gray boxes) and human (stippled boxes) forms for seven such anatomical markers. The appearance of archaic humans seems to coincide with an abrupt transition from a primate-like to the modern human form in all seven of the anatomical markers of

1. “Just intonation” is derived from tuning achieved between practiced singers adopting harmonic series relationships, rather than in answer to the referent of a keyboard tuning in equal temperament.

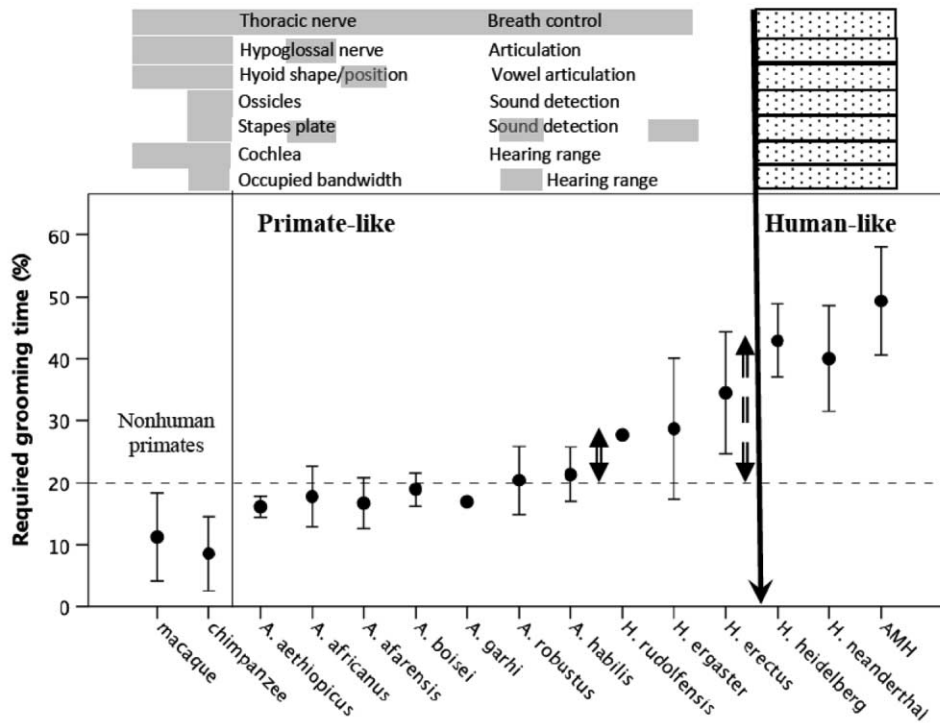


Figure 4. Mean (± 2 SD) for estimated grooming time requirement for the main hominin fossil species, compared to macaques and chimpanzees (for how these are calculated, see app. S6). Values for Neanderthals are corrected for their larger visual system (Pearce, Stringer, and Dunbar 2013); those for *Homo erectus* are uncorrected. No primate species devotes more than 20% of its day to grooming (horizontal dashed line; Dunbar 2022; Lehmann, Korstjens, and Dunbar 2007). Double-headed vertical arrows indicate two major phase shifts in required grooming time for which speech and/or singing could have evolved to fill a significant bonding gap. Mapped across the top are the presence of primate-like (gray) and humanlike (stippled) anatomical indexes of speech where these are known for individual taxa (with blank spaces indicating no data available). The seven anatomical indexes are indicated on the left and their functions on the right of each line. The bold vertical arrow marks the point where all the indexes seem to switch together. AMH = Anatomically Modern Human (*Homo sapiens*) fossils. For sources for anatomical data see appendix S6.

vocal production and detection. The change in these markers seems to signal a new importance for a vocal channel that is very different from the conventional vocal communication of monkeys and apes. It is unlikely that this vocal channel involved fully modern language since archaics did not have the cognitive mentalizing capacity to support fully fledged human language (Dunbar 2009, 2016; Pearce et al. 2014). This is, hence, more likely to signal the appearance of singing in the form of chorusing. In other words, chorusing (wordless singing, or singing with limited verbal content) is most likely to have evolved around 500,000 years ago and functioned as an important supplementary bonding mechanism for archaic humans (see also Dunbar 2012a; Mithen 2005).

An Integrated Model

Our overview of the likely functions of singing yielded prima facie evidence in support of five different phenomena: singing as a mechanism for communal bonding, OE, mate choice/retention, mother-infant crooning, and male bonding. This makes all of them correlates of singing but not necessarily their causes. Statistically, one of these will have greater impact on

fitness than the others, and so be the original cause, with the others being either secondary selection factors that reinforce the effect of the primary factor (but arise only because of the primary cause) or the consequence of windows of evolutionary opportunity (exaptations). In order to determine which of these is which, we need to determine the likely order in which the five elements appeared. There are 120 different possible sequences in which these five elements can be ordered. Evaluating all of these is beyond the scope of this paper, so we first seek to reduce the number of options by identifying the earliest likely dates for them. Figure 4 sets the benchmark by identifying 500 ka as a likely date for the evolution of the ability to articulate (and hear) sounds with sufficient precision to sing. Whatever function triggered the evolution of anatomical adaptations for vocal complexity (speech, but not necessarily language) at this point must thus be the original selection factor. Figure 4 specifically identifies the need to find behaviors that supplemented grooming for bonding large communities as an issue.

The other function that should be easy to date is mother-infant crooning (Falk 2004). In the light of contemporary human behavior in all cultures, this almost certainly evolved in combination with rocking to sooth grizzling infants. Rhythmic

rocking is known to sooth infants (Gursul et al. 2018; Korner and Thoman 1972), most likely because vestibular and/or cochlea stimulation caused by rocking triggers the endorphin system via spiral ganglion neurons whose sole function is to respond to head movements (Dunbar et al. 2021). Mothers singing to their infants might well have enhanced this endorphin effect because singing automatically entrains the rhythm of the rocking. Since monkey and ape infants do not exhibit the same level of grizzling (unlike tantrum behavior, which has a very different form and does not attract a soothing response in either humans or other primates), the timing for this is likely to be the point at which human babies first began to have an extended period of helplessness through infancy and early childhood (Morley 2017). Human lachrymose crying—an honest visual signal of the need for comfort not exhibited by chimpanzees—might have reinforced this (Trimble 2014). If a rocking-and-crooning response is a response to very young, helpless babies, its origins would be at the time when reproduction switched from a generalized primate-like schedule to a modern human one in which babies are born approximately 12 months premature compared to their ape equivalents (Martin 1990; Rosenberg 1992). This transition appears to have a late origin, associated with the appearance of anatomically modern humans ca. 250 ka (Gruss and Schmitt 2015).

Dissanayake (2008) has argued that this kind of rocking and crooning to calm the infant became co-opted to permit the emergence of intentional vocalization and response to partner affection available to both genders. Later, males may have started joining in as part of the process of pair bonding with a mate. If so, it must postdate mother-infant crooning and so also be associated with anatomically modern humans. This would fit with the anatomical evidence suggesting that a mating system based on pair-bonded monogamy was not characteristic of either early *Homo* or archaic humans and appears only with anatomically modern humans (Nelson et al. 2011; Shultz et al. 2014).

So far, then, we have a sequence that begins with communal chorusing and much later gives rise to maternal crooning, leading to crooning or singing as a possible mate choice and/or mate bonding mechanism. This leaves us with two final features to slot in: OE and male bonding.

It is implausible that the anatomical mechanisms required for male bonding would have evolved before communal bonding, since that would require males to have evolved a feature that females independently evolved only later. A common feature is likely to have evolved simultaneously in both sexes. This implies that the use of singing to bond males necessarily postdated the use of singing to bond groups. Whether it predates or postdates maternal crooning and/or pair bonding depends on whether male coalitions (most likely for raiding and defense against raiders) evolved early or late. It is not relevant to modern human hunting, since, ethnographically and archaeologically, this typically involved bow-and-arrow “stealth” hunting that is best done alone or in very small groups of two or three individuals at most. A more plausible case could be made for Neanderthals

who engaged in spear-based “confrontational” hunting in which success depends on the (presumably male) hunters being sufficiently bonded to stand together in the face of very large-bodied dangerous prey (Gaudzinski-Windheuser et al. 2018). In modern humans, group hunting occurs only in the context of lion hunts by young warriors of the Nilo-Hamitic tribes of northeast Africa (e.g., Maasai, Galla) or raids against neighboring tribes (the predominant form of warfare prior to settlement in villages after 12 ka); neither of these normally involve(d) singing but are invariably initiated and executed in silence and considerable secrecy, often aided by drugs (Lehmann and Mihalyi 1982). Singing and dancing occurred in the celebrations on completion of an expedition or in men’s groups as social contexts but never as the prelude to such an expedition. Since other archaic humans (e.g., Heidelberg) did not appear to engage in confrontational hunting to the same extent, this would imply that this function arose after 350 ka with the appearance of the Neanderthals as a distinct species in Europe, and that would necessitate a separate mutational event if this was also to be true for the anatomically modern human lineage. On the whole, that would seem implausible.

In contrast, if singing’s original function was as a form of chorusing for group bonding, then OE can probably be considered to have the same early origins as singing since there would likely be intense selection for harmonic convergence. This would have been necessary because all extinct fossil hominins, especially after the appearance of *Homo*, were sexually dimorphic in body size and the two sexes’ vocal ranges would inevitably have been very different.

Taken together, this would suggest the following time course: chorusing as a social bonding mechanism evolved around 500 ka with the appearance of archaic humans, followed very soon afterward by the evolution of OE; maternal crooning may not have evolved until around 300,000 years later with the emergence of anatomically modern humans, followed soon afterward by the use of singing or voice matching in mate choice and retention contexts. The use of singing for male bonding is less clear cut; it might have evolved earlier among the archaic humans but was perhaps more likely to have emerged with anatomically modern humans since the larger communities characteristic of this taxon will have had more males to bond than the smaller communities characteristic of archaic humans.

Challenges

Our aim has been to offer an explanation for OE, rather than for singing or music making in general. We here briefly consider some potential challenges to the proposal we make. We will not consider Pinker’s (1997) assertion that music has no adaptive significance but is instead simply a form of “auditory cheesecake” that excites responses in neural substrates that developed for other purposes: this now seems irrelevant in the light of recent evidence that music does have an evolutionary function. A more serious alternative explanation, derivative of G. Miller’s

(2000a) proposal that musicality evolved as a sexually selected trait, is that OE evolved to support mate choice and, perhaps, mate retention. The strongest evidence against this claim is that pair bonding does not seem to have evolved before anatomically modern humans, whereas the capacity to sing is demonstrably considerably older.

An alternative challenge arises from the argument that musical preferences are shaped more by culture than by innate responses (List 1984) and hence that features such as OE are not human universals. This view has been most forcefully expressed in a series of studies of the Amazonian Tsimane' people. McDermott et al. (2016) studied the differences in musical perception (expressed as preferences for consonant and dissonant chords) between urban listeners in the US and Bolivia compared to the Tsimane' and used this to reject the claim that the discrimination of harmonic is a human universal. They argued, instead, that harmonic is shaped by media experience not available to the Tsimane'. In a follow-up study intended to test for OE through requiring vocal imitation of synthesized sounds in a high range, Jacoby et al. (2019) found that Tsimane' tended to reproduce the interval of the fragment, but not in a manner that conveyed octave displacement. This is itself significant, however, since it suggests that the Tsimane' were executing a transposition of the stimulus, something that is actually more complex than simple octave mapping. The implication is that Tsimane' do in fact have a musical talent that allows them to preserve melodic frequency ratios rather than simply reproduce pitch chroma. In addition, Bowling et al. (2017b) have pointed out that the Tsimane' utilize the pentatonic scale in their own music and also doubted whether preference ratings represent the most suitable way of evaluating abilities. More importantly, perhaps, no report has been made of how these musical phenomena are transmitted from adults to children or, crucially, whether men and women can sing the same songs an octave apart. In short, it would seem that these studies are largely inapplicable to our present concerns.

Nonetheless, it may well be true, as Savage et al. (2021) have argued, that musicality arose as a consequence of some form of gene-culture coevolution—although, in fact, this is probably more correctly described as the well-known Baldwin effect (Baldwin 1896) whereby behavior leads genetic change (Podlipniak 2017). However, this suggestion addresses the inheritance mechanisms that underpin evolutionary dynamics and tells us nothing about the fitness functions that music or singing subserves. In contrast, our proposal is concerned only with the trait's evolutionary function. As is well understood, function, mechanism, and inheritance are independent components of Tinbergen's (1963) "Four Whys," and, as Tinbergen pointed out, the answer to one has no implications for the answer to others.

Future Research

This paper aimed to provide an evolutionary explanation for the phenomenon of OE and the reasons why men's voices

should be so much deeper than would be expected for body size compared to the voices of children and women. Future research that would illuminate the questions addressed includes the following:

1. The employment of Trần Quang Hai's model (Trần Quang and Bannan 2012; app. S1) of the dimorphic mapping of the four registers of the human voice (see also Henrich 2006; D. Miller 2000), testing a statistically significant number of child and adult human subjects in pairs as well as individually in response to a "neutral" stimulus to ascertain the pitch and timbre of the most common responses cross-culturally when male and female voices interact and when adult males interact vocally with children.

2. Acoustic analysis of the actual pitch of overtone sequences elicited by different modes of pitched performance, compared through same/different listener ratings. Such a project would help to discern the features of vocal production to which listeners respond, whether it be the distinct pitch (the partial in relation to a fundamental) or a more general timbral perception. This kind of study could help to illustrate the frontier in perception between the assumed conditions of music and speech and the common properties of both that evolved together as a consequence of human anatomy and auditory perception.

3. Listener ratings of vocal interaction with a measurable independent sound source that permits interval preferences to be elicited in practice, adopting an experimental method based on the studies of Arom (2010) and his laboratory.

4. Ethnomusicological evidence of how children learn songs from male and female models and how paired and group singing illustrate patterns of harmonic interaction and their global distribution.

5. Quality and salience of listener and performer responses to the differences in terms of physical and hormonal reactions to measurably well-tuned and blended vocal unisons and harmonic reinforcements. Since group vocalization triggers the endorphin system, we would expect endorphin activation to be higher when singing is in well-tuned unison than when it is less accurate in its pitching and lacks the reinforcement of harmonic coherence and to exhibit a similar peak in endorphin output when listening to men and women singing in OE.

6. Performer self-report and group report of the social-bonding effects of group vocal participation, including more direct measures of the extent to which singers feel bonded to each other when singing in and out of melodic unison (both 1:1 and the 2:1 unison OE).

7. Some simple tests of the mate choice/maintenance hypothesis would be (i) to compare the voices of established couples and (ii) to ask individuals to rate the pleasantness/unpleasantness of a set of target voices when they speak or sing with them.

8. The five key functions (motherese, mate choice, courtship, male bonding, communal bonding) could have emerged in 120 different possible orders: finding evidence to identify the correct order will provide deep insights into their relative importance in terms of fitness.

Conclusions

OE is the perceptual-productive mechanism, genetically endowed and released through vocal practice and imitation, that activates and refines the feedback loop between auditory and vocal processes during singing. It may have evolved from two competing pressures on males: the demands of sexual selection so as to exaggerate size in male-male competition through lower vocalization frequency and the need for harmonization as a social-bonding mechanism (both within groups and, perhaps, between sexual partners). The arrival at the octave represents a compromise that allows male voices to be lower in frequency while still acoustically complementing female and juvenile voices. It permits unison singing, which promotes social bonding, enhances group coordination, and may support mate retention through paired activity (Dissanayake 2008; Weusthoff, Baucom, and Hahlweg 2013) as well as social bonding more widely but also presents the best acoustic framework for vowel matching that relates to adaptive human processing of the harmonic series as the determinant of social listening. Human capacity to match vowels in language learning depends on this, as do other interpretants of timbre—detecting age, health, state of mind, preparedness to interact (acoustic contagion), and level of perceived physical investment in communication (Bannan 2020; Snowdon, Zimmermann, and Altenmüller 2015). These abilities indicate a role for pitch sensitivity in theory of mind (Baimel et al. 2015; Livingstone and Thompson 2009) that is also consistent with social brain theory (Dunbar 2003; Savage et al. 2021). While these factors would inform natural selection and sexual selection models focused on reproductive success and survival, the plasticity of these sophisticated abilities also contributes to cultural manifestations such as ritual, play, and the marking of social milestones on which quality of life may also be built. The communicative Rubicon was crossed when male voices reached a range at which they were as different as possible from those of children and adult females while being perceived acoustically as being “the same,” with all the benefits for musical and linguistic generativity that this confers.

Comments

Alan R. Harvey

School of Human Sciences and Conservatorium of Music, University of Western Australia, Western Australia 6009, Australia, and Perron Institute for Neurological and Translational Science, Western Australia 6009, Australia. 11 X 23

The paper by Bannan, Dunbar, and Bamford addresses an interesting question about human musicality and communication: why, after puberty, is the male voice approximately one octave lower in pitch than the adult female voice? They convincingly argue that this difference is not what would be expected based on body size or “generalized dimorphism.”

While it has been argued that deeper male vocalizations in many primates are especially important in larger groups and are associated primarily with sexual selection (Aung et al. 2023), Bannan, Dunbar, and Bamford propose that octave equivalence in humans is an adaptation that allows “chorusing for social bonding and pitch matching in intergenerational exchange”—did the authors also mean *intragenerational* exchange here? It is not clear how exact this equivalence needs to be; thus, the authors consistently claim that the *average* male-female voice difference is *exactly* one octave (my emphasis) but sometimes describe this difference as “precise” or “approximate” (see also Jaatinen et al. 2023). Could any interindividual variability in the perception of pitch chroma be important from a selection or social interaction perspective?

Some birds have been shown to move to a beat, and many animal species engage in to-and-fro vocalizations, but the abilities to sing and move in synchrony with others and to match pitch—whether in unison or in harmony—are special sensorimotor skills of *Homo sapiens* (Della Bella, Berkowska, and Sowiński 2015; Harvey 2017). In this regard, while there is evidence of octave generalization and similar processing of harmonics in at least some primates (Wright et al. 2000), imaging studies on humans with what some term “congenital amusia” reveal a retained ability to discriminate pitch (Moreau, Jolicœur, and Peretz 2013) but discrete connectional abnormalities linking sensory auditory processing to the frontal cortex (Hyde et al. 2006, 2007), suggesting a deficit in human auditory-motor feedback systems (Mandell, Schulze, and Schlaug 2007). Anticipation and synchronized entrainment of rhythm are also critical in a group context (McPherson et al. 2018; Repp et al. 2013). In their paper, Bannan, Dunbar, and Bamford rightly focus on the importance of these coordinated human sensorimotor behaviors for enhancing cooperativity, empathy, and social bonding, as has been widely discussed by many scholars (e.g., Cross 2003; Dunbar et al. 2012; Harvey 2017; Merker 2021; Mithen 2005; Patel 2008; Savage et al. 2021; Storr 1992; Tarr et al. 2015). However, there are several aspects of their discussion that are open to alternative perspectives.

For example, based on fossil records they speculate that archaic human “singing” is of the order of 500,000 years old and that such communal activities were necessary to replace “conventional” social grooming as groups got larger. This is suggested to be a form of nonverbal chorusing, although the authors admit there may have been some “limited verbal content.” In this regard, gossip has also been put forward as a driver of social cohesion (Dunbar 1996). The authors also consider the development of communication between caregivers and preverbal offspring (crooning or motherese). I have argued elsewhere (Harvey 2017), similar to suggestions by others (Brown 2000; Thompson, Marin, and Stewart 2012), that the term “musilanguage,” or “protolanguage,” is best used to describe the precursor from which both music and language later evolved. Similarly, careful evaluation of the neurobiology of motherese suggests that it is neither music nor language but perhaps more akin to the musilanguage communication system

used by archaic humans (Harvey 2017). Anthony Storr (1992) wrote that “we can perceive that language and music were originally closely joined, and that it makes sense to think of music as deriving from a subjective, emotional need for communication with other human beings which is prior to the need for conveying objective information or exchanging ideas.” But music as understood by cognitively modern humans is not propositional and poses no testable hypotheses; it “neither ploughs, sows, weaves nor feeds” (Cross 2001). It seems to me unlikely that the early evolution of hominin communication contained *no* useful learning content or information transfer; thus, I prefer to use a term such as “musilanguage” and limit the term “music” to modern humans in order to differentiate it from the other universal communication stream, language and articulate speech, an evolutionary transition that likely occurred within the past 100,000 years (Harvey 2017).

There is of course much useful information that can be gleaned from careful study of hominin fossils, but, as I have argued at length (Harvey 2017), cranial casts provide only a guide to the possible cognitive and communicative capability of our ancestors. They cannot tell us about the underlying neural architecture, connectivity, or plasticity, let alone the size or exact function of a given region (Van Essen et al. 2012): “it is really not possible to establish with certainty what side a language-related area is located in a human brain merely by gross examination of the external anatomy, so to try to do so in ancestral fossil casts is speculative to say the least, if not impossible” (Harvey 2017). In humans, for example, compared to monkeys there are far more connections between the laryngeal motor cortex and the parietal sensory cortices, thus permitting greater sensorimotor modulation and integration (Kumar, Crosson, and Simonyan 2016). There are also unique projections from the phonatory region of motor cortex to brain stem nuclei that control vocalization (Zarate 2013), neural specializations that are not obviously reflected in a morphological analysis of cranial nerve foramina. As one further example, the human orbit varies only a little in size (Bekerman, Gottlieb, and Vaiman 2014), yet imaging of the human visual cortex reveals a 2–2.5-fold variation in location and surface area (Dougherty et al. 2003), with the size of the primary visual cortex not scaling with overall brain size (Schwarzkopf, Song, and Rees 2011).

Bannan, Dunbar, and Bamford discuss the impact that communal music making and maternal-infant interactions have on the endorphin system (usually measured using the surrogate of the raised pain threshold). While the endorphins are clearly of great importance, mention perhaps also needs to be made of the role of dopamine in expectation, reward, and entrainment and the critical importance of oxytocin in maternal-infant attachment and in helping to drive prosocial interactions, empathy, and within-group cooperation (Harvey 2020). Oxytocin not just works in an endocrine fashion but also acts as an important neuromodulator within many regions in the central and peripheral nervous systems, and at the receptor and synaptic levels the peptide can interact with both the dopaminergic and

opioid systems (Paletta et al. 2022; Putnam and Chang 2022). Its role in attachment and bonding is clear (e.g., Fujiwara et al. 2019; Kim et al. 2014; Seltzer, Ziegler, and Pollak 2010; Strathearn et al. 2009), and in adults measured oxytocin levels are higher with communal music making (Grape et al. 2003; Keeler et al. 2015; Kreutz 2014; Ooishi et al. 2017).

In summary, there have been a number of studies addressing octave equivalence in humans, and Bannan, Dunbar, and Bamford make an important new contribution to this discussion. Perceptual fusion of musical notes, especially for the octave, and recognition of dissonance and thus the representation of harmonicity, appears to be cross-cultural and a human universal (Demany et al. 2021; McPherson et al. 2020), although there is evidence that twenty-first-century humans subjectively prefer octave intervals slightly above the 2:1 frequency, with this preference resolved by late-stage auditory processing (Jaatinen et al. 2023). In processing sound in the cochlea, compared to other species, including the primates, there are differences in cochlear tuning in humans—differences thought to emphasize spectral cues (Shofner 2023; Walker et al. 2019). However, some conserved elements remain, including the presence of nonlinear processing and harmonic distortions within the mammalian cochlea, involving complex mechanisms within the organ of Corti (Cooper 1998; He and Ren 2021). One important product of these distortions is the generation of harmonic frequencies above the fundamental f_0 , the largest and most prominent being $2f_0$, the octave directly above. Could the octave equivalence described by Bannan, Dunbar, and Bamford reflect this peripheral phenomenon in some way—the second harmonic distortion of the male voice as perceived by females being perceptually fused with their own f_0 ? Related to this, and given that the human auditory cortex processes both frequency and pitch (Allen et al. 2022), it would be of interest to image the brains of boys before and after puberty to determine and compare how neural processing changes in an individual as the voice descends.

Manuela M. Marin and Bruno Gingras

Department of Cognition, Emotion and Methods in Psychology, University of Vienna, 1010 Vienna, Austria (manuela.marin@fh-krems.ac.at), and Austrian Research Institute of Empirical Aesthetics, 6020 Innsbruck, Austria/Austrian Research Institute of Empirical Aesthetics, 6020 Innsbruck, Austria, and Faculty of Life Sciences, Department of Cognitive Biology, University of Vienna, 1010 Vienna, Austria (brunogingras@gmail.com). 19 IX 23

It's All in the Sound and in the Brain

While we appreciate the interdisciplinary efforts to investigate the evolution of a complex psychoacoustical phenomenon such as octave equivalence (OE), we were dismayed to see that, throughout the article by Bannan, Dunbar, and Bamford, the term was not defined clearly and sometimes used improperly. For example, in the first sentence of the abstract, the authors define OE as a phenomenon “whereby adult male voices are, on

average, an octave lower in pitch than those of adult females and children.” Later on in the abstract, OE is defined as the “capacity to share and respond to vocal pitches . . . as if they are ‘the same’ irrespective of the difference in range.” Then, in the introduction, OE is defined as “the ability to recognize that two notes are in harmony when an octave apart.” In the section “Octave Equivalence: Ontogeny and Phylogeny,” OE is very broadly defined as an umbrella term comprising three putatively related phenomena: “the perception of OE; the ability to produce octaves vocally, in imitation as well as simultaneous performance; and the average octave difference between male and female voices.” However, the established definition of OE is a perceptual one, namely, the perceived similarity of pitches separated by one octave, corresponding to a 2:1 frequency ratio (Burns 1999). It is legitimate to present a novel, more comprehensive definition of OE, but the justification for such a reappraisal requires a rigorous use of the existing terminology and definitions.

Taking Burns’s (1999) definition as a starting point, two tones separated by an octave are perceived as belonging to the same pitch class, although they differ in pitch height. Bannan, Dunbar, and Bamford claim that “there appears to be no obvious explanation for” this phenomenon. However, taking a psychoacoustic perspective, Helmholtz (1862) already explains in his book *On the Sensations of Tones* that the degree of perceived pitch relatedness (or similarity; in the original German *Tonverwandtschaft*) of two complex tones depends on the number of overlapping partials (harmonics). In the case of the octave interval, the perceived similarity is the highest outside the unison because the partials of the higher tone overlap with every second (even-numbered) partial of the lower tone. It is important to note that a doubling of the frequency of a complex tone does not lead to the perception that the two tones are the *same*. The pitch class is the same but not the pitch height, which is acknowledged in models of human pitch perception (Shepard 1982; see also Warren et al. 2003). The perfect fifth (3:2) and perfect fourth (4:3) are other intervals in which the harmonics overlap considerably, and together with the octave these intervals are important constituents of musical scales around the world (Davies and Ravasio 2021; for a link between musical scale preferences and the harmonic series, see also Gill and Purves 2009).

We argue that it is possible to explain the evolution of perceptual OE without considering group singing, group bonding, or the average octave difference between male and female voices in the first place. To be sure, this does not rule out the possibility that the sexual dimorphism of voice pitch may have evolved much later through group singing as a consequence of the already-existing phenomenon of perceptual OE among humans (see also Wagner et al. 2023). However, even using (one of) Bannan, Dunbar, and Bamford’s definitions of OE, it should be noted that the voice pitch difference between the human sexes is not exactly one octave, as they suggest, but somewhat less (e.g., 1:1.8 for speaking F0 [Pisanski et al. 2014], 1:1.9 for tennis grunts [Raine, Pisanski, and Reby 2017]).

Moreover, formants are better predictors of body size in humans than fundamental frequency (for a review see Pisanski et al. 2014), which should be kept in mind when relating perceptual OE to sexual dimorphism in voice pitch more generally.

The evolution of musical pitch structure, of which OE is one determining feature, may be rooted in the nature of the human auditory system and how sounds are processed in the brain (Trainor 2015). The auditory system decomposes incoming sounds into their frequency components and integrates this information into a pitch percept in the cortex by means of auditory scene analysis, as well as spectral and temporal frequency coding of the basilar membrane. The mammalian auditory system has evolved over millions of years to detect the nature and location of sounds in the environment (not just human vocalizations), with a unique middle and inner ear structure (Luo and Manley 2020) that enables spectral analysis of “rich” sounds with many overtones. This ability is, in our view, a basic prerequisite for the perception of OE, since otherwise the spectral similarities of two tones one octave apart would be undetectable, which means that the octave similarity percept, so crucial to music cognition, would likely be absent. We thus argue that the ability to perceive OE has probably preceded any kind of active imitation of speech or singing during phylogeny because OE is largely built on general sound-processing skills (see also Wagner et al. [2023], who report no link between OE and vocal learning). Likewise, in the ontogeny of speech and music, perception usually precedes production (for the case of OE see Demany and Armand 1984), and within this context one could argue that the perception of OE may act as a “bootstrap” (Pinker 1984) that facilitates the imitation of adult speech (Peter et al. 2015) by children who have a different vocal range. Bannan, Dunbar, and Bamford also mention that the “harmonic series relationships have shaped human hearing and aural engagement” (see also Bowling, Purves, and Gill 2018). Human vocalizations prior to speech and singing (e.g., musical protolanguage; Darwin 1871) must have been already “rich” and contained a high number of harmonics because otherwise speech, in which formants play a crucial role, could probably not have evolved. The potential role of harmonics in vocalizations for OE has also been identified by Wagner and Hoeschele (2022) by looking at the “harmonic clarity” in animal vocalizations.

Trainor (2015:4) argues that “music conformed to the human auditory system, rather than the other way round,” referring also to preexisting features of the brain. Indeed, a recent study by Pankovski and Pankovska (2022) has shown that a preference for a discrete set of musical intervals, as well as a preference for consonant intervals that match Helmholtz’s (1862) list of preferred intervals, emerges spontaneously when training a nonspecific Hebbian neural network with 12,000 complex tones spanning 10 octaves (with adjacent tones separated by 0.01 semitone, or 1 cent), even without OE mapping. This is in line with the idea that the origins of perceptual OE are mostly rooted in the physics of complex tones and the nature of the human auditory system.

Patrick E. Savage and Yuto Ozaki

School of Psychology, University of Auckland, Auckland, New Zealand, and Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan (patrick.savage@auckland.ac.nz)/Graduate School of Media and Governance, Keio University, Fujisawa, Japan. 19 X 23

Globally, Men and Women Do Not Sing and Speak Exactly One Octave Apart

We were pleased to read Bannan, Dunbar, and Bamford's interesting and timely target article arguing for an evolutionary hypothesis for the underexplored phenomenon of octave equivalence. The article nicely extends and synthesizes previous proposals (e.g., Dunbar 2012*b*; Savage et al. 2021; Wagner and Hoeschele 2022), including our own music and social bonding hypothesis. As such, it may not be surprising that we agree with their basic argument that octave equivalence facilitates social bonding. But we particularly welcome the way in which they outline the potential trade-offs between social bonding and sexual selection to arrive at a "compromise that allows male voices to be lower in frequency, while still acoustically complementing female and juvenile voices." Given that the presence or absence of sexual selection for human musicality has been historically controversial (Bertolo et al. 2023; Darwin 1871; Fitch 2006; Mehr et al. 2021; Miller 2000*b*), it is nice to see a nuanced proposal for how sexual selection and social bonding may have jointly shaped the evolution of human musicality.

That said, we felt the authors could have devoted more discussion to another major topic of long-standing debate: the evolutionary relationship between music and language—and particularly between song and speech. In this context, it is important to note that the key data point cited by Bannan,

Dunbar, and Bamford in support of their central claim that differences between the adult male voice and the adult female voice are "exactly one octave" is not from the singing voice but from the speaking voice. Specifically, Puts et al. (2016) analyzed recordings of 15 male and 15 female Hadza foragers from Tanzania speaking "hujambo" (hello) and 15 male and 15 female Michigan State University students from the US reading the first paragraph of the Rainbow Passage, finding mean vocal fundamental frequency (F₀) to be 115.3 Hz in males versus 224.6 Hz in females, or a male-to-female mean F₀ ratio of 0.513 (see table 1). However, F₀ can vary greatly depending not only on biological sex but also on features such as age and vocalization context, with Ozaki et al. (2022) finding in a global sample of 75 diverse linguistic varieties that singing is consistently much higher than speaking (by approximately half an octave).

Delving further into the data, we start to question Bannan, Dunbar, and Bamford's insistence that "the fact that the difference is exactly one octave is too precise to be an accident." While we do not question their broader point that the dimorphism is much larger than for other apes, 0.513 is technically not "exactly" 0.5, and Ozaki et al.'s singing data give an estimated ratio (0.556) that is still further from an exact octave. Indeed, comparison with other global singing and speaking data from various sources all produce estimates of F₀ ratios greater than 0.55 (table 1). Even the larger dataset from which the 0.51 estimate sample was drawn from (Puts, Apicella, and Cárdenas 2012) provides estimates of greater than 0.55 for both the US and Hadza (Tanzanian forager) subsamples. Other studies with even larger sample sizes are even further from an octave—for example, Berg et al.'s (2017) data from 2,472 German participants give a male-to-female F₀ ratio of 0.66. This all suggests that the sexual dimorphism in human vocal production does not quite align precisely with the octave.

Table 1. Comparison of previously published audio recordings of singing and speaking around the world

Dataset	<i>n</i> (male)	<i>n</i> (female)	Modality	Mean male F ₀ (Hz)	Mean male-to-female F ₀ ratio
Ozaki et al. 2022 ^a	41	34	Song	161.3	.556
Hilton et al. 2022 ^a	25	184	Song	152.6	.607
Ozaki et al. 2022 ^a	41	34	Speech	114.2	.571
Hilton et al. 2022 ^a	25	184	Speech	130.7	.623
Puts, Apicella, and Cárdenas 2012 ^b	176	268	Speech	111.4	.552
Puts, Apicella, and Cárdenas 2012 ^c	32	43	Speech	114.9	.551
Puts et al. 2016 ^d	30	30	Speech	115.3	.513
Pisanski, Groyecka-Bernard, and Sorokowski 2021 ^e	71	83	Speech	120.5	.583
Berg et al. 2017 ^f	1,154	1,318	Speech	111.9	.664

Note. This table showing the comparison of previously published audio recordings of singing and speaking around the world suggests that the adult male-to-female vocal fundamental frequency (F₀) ratio is greater than 0.55 (~1,000 cents/minor seventh), not exactly 0.5 (1,200 cents/one octave), as claimed by Bannan, Dunbar, and Bamford. Note that these data include only adult-directed singing/speaking (infant-directed song/speech from Hilton et al. [2022] are excluded, as they showed them to be substantially higher pitched than adult-directed song/speech).

^a Global.

^b Study 1, United States.

^c Study 2, Hadza, Tanzania.

^d *n* = 60 balanced subset of Puts, Apicella, and Cárdenas (2012) with equal numbers of Hadza and US men and women (cited by fig. 1).

^e Poland; taken from Free Speech data.

^f Germany; taken from the level II data of all age groups.

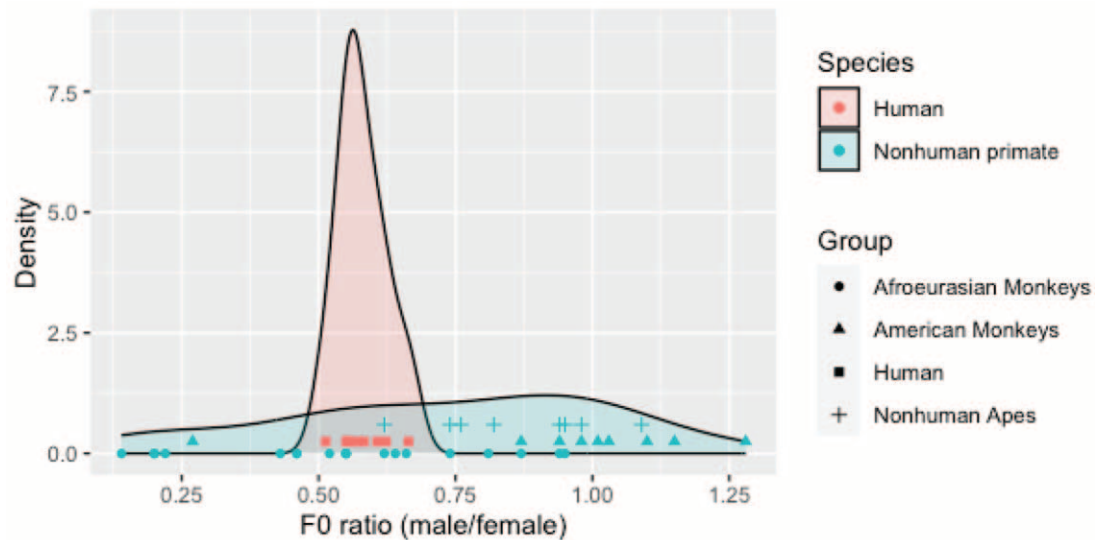


Figure 5. Male-to-female fundamental frequency (F_0) ratios of human data in table 1 and primate data taken from Aung et al. (2023).

This does not necessarily undercut Bannan, Dunbar, and Bamford’s argument—the dimorphism is indeed *almost* an octave, and this is much greater than for other apes (Aung et al. 2023). Since human vocal pitch production is quite flexible, it is clearly close enough to an octave to enable men and women/children around the world to sing in octave harmony (Brown and Jordania 2013; Mehr et al. 2019; Savage et al. 2015), which is indeed generally perceived as unison via the phenomenon of octave equivalence. However, when we examine the F_0 ratios of primates in general, it becomes apparent that some Afroeurasian monkeys also have a near-octave difference (fig. 5). This suggests the need for further cross-species and cross-domain research to investigate potential interactions between spoken and sung vocal pitch and octave equivalence in human and nonhuman animals.

If singing did indeed evolve before speaking, as proposed by Bannan, Dunbar, and Bamford, why do both men and women vocalize at a substantially lower pitch when they speak than when they sing? There may be more trade-offs than simply sexual selection and social bonding involved. For example, communication pressures may have led vocal production to specialize for ranges to optimize perceived fundamental frequency for melodic discrimination in song versus higher harmonics for vowel/consonant discrimination in speech (Ozaki et al. 2022; Patel 2008). Additionally, voice pitch may be related to the formation of social bonding, as high-pitched voices are perceived as more trustworthy, cooperative, and warmer than low-pitched voices (Knowles and Little 2016; O’Connor and Barclay 2017; Wu et al. 2023; though cf. Schild et al. 2020).

None of this necessarily detracts from the value of Bannan, Dunbar, and Bamford’s hypothesis. Instead, it highlights the complex nature of the evolution of vocal communication and the need for more research along the lines outlined in their “Future Research” section. We look forward to seeing the results.

Acknowledgments

We thank Joshua Bamford, David Puts, and Courtney Hilton for providing feedback on draft versions of the manuscript. We are supported by funding from the New Zealand government, administered by the Royal Society Te Apārangi (Rutherford Discovery Fellowship 22-UOA-040 and Marsden Fast-Start Grant 22-UOA-052), and from the Japanese government, administered by the Japan Society for the Promotion of Science (KAKENHI Grant-in-Aid 19KK0064).

Data Availability

All data in table 1 and figure 5 are taken from the publicly available sources in the relevant publications (Aung et al. 2023; Hilton et al. 2022; Ozaki et al. 2022; Puts, Apicella, and Cárdenas 2012, 2016).

Hollis Taylor

Sydney Conservatorium of Music, University of Sydney, 1 Conservatorium Road, Sydney, New South Wales 2000, Australia (hollis.taylor@sydney.edu.au). 21 VIII 23

Octave Equivalence: A Perspective from Zoömusicology

This article draws attention to an intriguing but vexing phenomenon: the significant role that the interval of an octave plays in human pitch perception and singing. The extensive literature review above draws on disparate disciplines, reminding us that even the most basic questions vis-à-vis human music—“When did music evolve?” or “What is music for?”—lack straightforward answers. The authors perform an important function in wresting vocal dimorphism as found in octave equivalence (OE) from the

overly simplistic explanatory grip of sexual selection, insisting on a wider, deeper look at OE's functional and evolutionary origins in biological processes set in motion millions of years ago.

Given the complexities of and long-standing debates within evolutionary biology, it would be naive for a zoömusicologist to claim mastery of the topic. Thus, I will refrain from assessing the authors' ambitious effort to chronologically order key features of singing in order to date its origins. Instead, I will focus the bulk of my comments on the introduction, which itemizes protomusical features found in humans. The story told in the first paragraph moves too quickly for me, giving the impression that these behaviors are unique to humans, or likely so, or at minimum "unmatched in any other species." Assembling an account upon the sand of human exceptionalism is a problematic and undependable ground upon which to build.

Exceptionalism claims clash with scientific method and challenge Carl Sagan's tenet that absence of evidence is not evidence of absence. Since very few of the world's species have been examined, we lack sufficient knowledge to make such declarations. In this, I join de Waal (2016) in appealing for a pause on claims of human uniqueness. Additionally, while I appreciate that statements of (statistical) uniqueness aim to simplify this complex subject, I remain unconvinced that this is ultimately helpful. This journal's readership of anthropologists and interested generalists stands to benefit from a fuller portrayal of this topic's messiness regarding animals: nagging exceptions, areas of controversy and contradiction, and deep pockets of remaining ignorance. I begin at the top of the authors' list:

Entrainment. Since evidence of beat synchronization and motor entrainment was first reported in animals in 2009, the list quickly expanded to three species of parrots, rats, chimpanzees, and a California sea lion, with ongoing investigations into horses. In addition, Albert's lyrebirds are known to sing and dance to their own strictly measured beat, which is underreported. As methods for testing and data analysis are refined and more species are studied, our cross-species understanding of entrainment will advance. "Unmatched" in humans? Perhaps, but this is currently provisional.

Pitch center recognition. The Podlipniak article upon which the authors base their claim of humans' "unique capacity to match pitch precisely" is itself vague. Animals do have difficulty recognizing transposed melodies, but, in another framing, birds surpass mammals (including us) in their capacity for absolute pitch. A rare trait among humans, absolute pitch could make transposed melodies sound wrong or different to birds. Perhaps with this ability, they have no need for a pitch center. Further, the pitch center concept (e.g., "the key of G major") is more pertinent to Western music than other genres. Most animals sing solo (a horizontal melody), not in ensemble, and solo song is not accompanied by chords that add vertical complexity and confirm a tonal center. Nonetheless, birds do contribute to a dawn chorus's vertical reality, which is "organized by social communication extending beyond species boundaries [that sees] increased stereotypy and synchrony among species" (Tobias et al. 2014:1020).

Space does not allow me to refute or expand upon each item. In any case, what is central to my thesis is that minimizing animal

abilities rather than researching and celebrating different forms of intelligence stands to encumber new knowledge. The authors conclude their list by stating, "Few of these behaviors are apparent in our closest primate relatives." However, such relatives lack the capacity for vocal learning so foundational in music. For this, we must turn to songbirds, parrots, hummingbirds, seals, dolphins, and whales, who share the rare capacity of vocal learning.

That humans seem to uniquely possess OE is attributed by the authors to Hoeschele, Weisman, and Sturdy (2012); however, I find in those pages an enthusiastic preparedness to investigate OE in songbirds. Add to this another complication: not all human cultures exhibit OE. Will (1997) finds no evidence for it in central Australian vocal music, which is sung in unison rather than the octave. He concludes, as have others, including Burns (1999), that OE is a cultural agreement.

Granted, human communal chorusing does occur at the octave, but largely absent from this article is a mention of other intervals. The case of the Amazonian Tsimané people is perhaps a placeholder for numerous instances of parallel singing not at the octave, but this is not made explicit. Sachs (1943) details how singing on the Andaman Islands and in Tierra del Fuego occurs in parallel fourths and fifths; so are traditional Icelandic songs. In a list that is far from comprehensive, singing in parallel seconds is known in Bulgaria and the Admiralty Islands, thirds in Hispanic culture, thirds and sixths in Croatia and Goa, and seconds and fifths in Jewish culture.

Without undercutting its conclusions, would anything in this well-rehearsed and exceedingly commendable effort to date the origins of singing and OE's role in it change if claims of human uniqueness were omitted? Animals are the product of the same evolutionary pressures as humans. As Darwin maintained, the differences between human and animal behavior are often of degree and not kind. Birds, for instance, have taken a different neural route from humans to generate complex cognition. To me, eschewing claims of human uniqueness will allow us to better test and understand the evolutionary roots of human musical capacities in comparative studies with animals.

**Bernhard Wagner, Daniel L. Bowling,
and Marisa Hoeschele**

Acoustics Research Institute of the Austrian Academy of Sciences, Wohllebengasse 12–14, 1040 Vienna, Austria (marisa.hoeschele@oeaw.ac.at) (Wagner and Hoeschele)/Department of Psychiatry and Behavioral Sciences, Stanford School of Medicine, MSLS Building, 1201 Welch Road, Stanford, California 94305, USA, and Center for Computer Research in Music and Acoustics, Stanford University (Bowling). 4 XII 23

Defining Octave Equivalence and a Comparative Approach to Understanding Its Biological Foundations

Bannan, Dunbar, and Bamford set out to explain the considerable difference in average voice pitch between human males

and females in terms of an evolutionary (adaptive) role for music in supporting social bonding among early hominids. They argue that while some degree of difference between male and female pitch can be expected on the basis of body size dimorphism (and the role of the voice in male-male competition; Dunn et al. 2015; Pfefferle and Fischer 2006), this factor alone cannot explain why male and female voices adhere to the particular pitch difference of 2:1 (an “octave”) when we sing together. Bannan, Dunbar, and Bamford propose that octave-separated singing—multiple individuals singing together in synchrony with a pitch difference equal to an octave—confers stronger social bonds on participants (men, women, and children), thereby increasing their fitness.

We are grateful for this effort to contextualize the octave, referred to as the “basic miracle of music” (Cooper 1973), within the framework of human biology. We also agree that an answer to the age-old problem of octave equivalence (OE) may well be found in the social functions of music, at least in part. Here, we aim to increase the impact of this work by offering a strict definition of OE and describing how current cross-species research seeks to inform biological foundations.

Bannan, Dunbar, and Bamford use multiple definitions of OE in their paper, variably describing it as (1) “a unique vocal property . . . whereby adult male voices are, on average, an octave lower in pitch than those of adults females and children”; (2) “a phenomenon in pitch perception [comprising] the ability to recognize that two notes are in harmony when an octave apart”; and (3) “the ability to produce octaves vocally.” This multiplicity of definitions for OE generates ambiguity about its precise nature. Moreover, it can lead to circular reasoning by treating the link between OE and voice pitch dimorphism as a given, rather than a hypothesis. A more tradi-

tional definition from music cognition specifies OE as “the assumption that tones separated by an octave are in many ways musically identical” (Burns 1999:252). While even this definition leaves something to be desired (musically identical?), the key point is that we tend to perceive tones separated by an octave as so similar that despite their marked difference in pitch, two people singing an octave are perceived as singing the same thing.

By disconnecting the perception of OE from the behavior of octave-separated singing, the above definition allows for at least two additional hypotheses, neither of which necessarily excludes that of Bannan, Dunbar, and Bamford. One such hypothesis derives from the fact that the octave interval occurs within any harmonic vocalization (see fig. 6), which is common in many species, including humans (Bowling et al. 2017a). This physical fact of the harmonic series means that the vast majority of our experience with octaves, whether on individual or evolutionary timescales, comes from single sources (Bowling and Purves 2015; Bowling, Purves, and Gill 2018; Bowling et al. 2017b). This may account for our propensity to perceive octave-separated singing as a single unified voice. An even more basic hypothesis derives from a purely physical standpoint: perhaps we perceive two harmonic sounds separated by an octave as the “same” because they are as similar as two harmonic sounds with different fundamental frequencies can possibly be, with overlap in acoustic frequency content at a maximum (see fig. 6).

Progress in evaluating hypotheses related to OE can be made through cross-species research. For example, while Bannan, Dunbar, and Bamford’s hypothesis implies that OE is intimately related to voice pitch dimorphism and chorusing, the first alternative described above implies that OE may instead (or additionally) arise from an evolved sensitivity to the

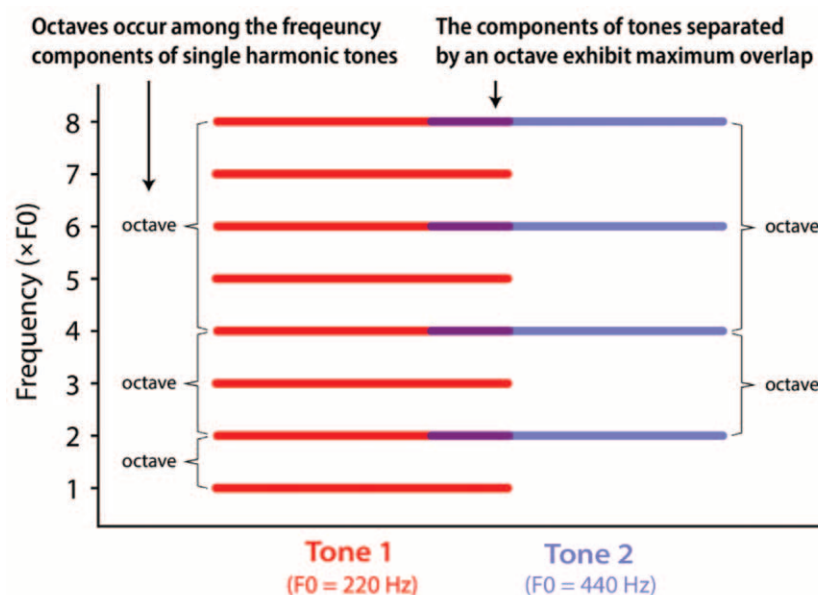


Figure 6. Overlapping harmonics of two sounds separated by an octave. Based on Wagner et al. (2022).

harmonic content of conspecific vocalizations. If so, OE may be present in other similarly sensitive species. Likewise, the physical hypothesis described above implies that all animals should perceive OE, at least to some extent (presuming they are sensitive to harmonics).

Beyond helping evaluate these possibilities, comparative study also allows us to test the role of underlying biological traits. Wagner and Hoeschele (2022) recently reviewed cross-species OE research, describing an approach to systematically determining its biological roots in terms of key traits related to auditory-vocal behavior. Specifically, they hypothesized that OE requires four key traits and described how their variable expression in specific species can be used for empirical testing. Two such traits, also discussed by Bannan, Dunbar, and Bamford, are sexual dimorphism in voice pitch and chorusing. A third is our heavy reliance on clear harmonic information in vocal communication, underpinned by the (presumably) adaptive loss of secondary vocal membranes in the human larynx (these complicate vocal production dynamics and obscure harmonic content in other primates; Nishimura et al. 2022). When in Bannan, Dunbar, and Bamford's time line might this loss have occurred? Was it a prerequisite for OE? The fourth trait indicated in Wagner and Hoeschele (2022) is vocal learning. When human children learn new speech and song, they need a way to accommodate pitch differences during imitation. Notably, children spontaneously use OE when imitating adult voices outside of their range (Peter, Larkin, and Stoel-Gammon 2009; Peter, Stoel-Gammon, and Kim 2008; Peter et al. 2015). In general, the smaller size of developing offspring makes voice pitch differences far more common as a function of age than sex, implying that testing juvenile animals for evidence of OE may generate more data.

In sum, the bridge built by Bannan, Dunbar, and Bamford between OE and human biology goes a long way toward rationalizing—a core feature of musical phenomenology—and we agree with much of their account that it is related to human sexual dimorphism in voice pitch. However, we also stress the importance of clear definitions and consideration of additional or alternative possibilities, proposing that comparative research offers an important path forward.

Reply

We thank the commentators on our article as well as their appreciation of its potential to elicit interest across the range of disciplines involved in this topic. We well recall the key influence of Livingstone's 1973 *CA* article "Did the Australopithecines Sing?" and the way in which commentators dissected its import, illustrating their own partialities as well as presenting perspectives on aspects of Livingstone's proposal that proved open to further investigation. We hope we can rise to the same challenge.

Harvey's response illustrates an important alignment to our case, with a couple of specific differences of emphasis. His preference for a "musilanguage" model over our focus on a Darwinian sequence (Darwin 1871:chap. 19) is shared by many but is in our opinion largely semantic: if a communication system does not (yet) possess grammar, syntax, or vocabulary, but comprises pitched and rhythmic content capable of being learned, remembered, varied, shared, and performed simultaneously, it is semiotically closer to music than to speech. We are perfectly content with that. Harvey rightly reminds us of the need to broaden the behavioral and motivational role of endocrine responses, though our focus on endorphins is based on the very specific characteristics of this family of neuroendocrines. Oxytocin has a very short half-life (measured in minutes) compared to endorphins (measured in hours; see Dunbar 2021); moreover, large-scale genetics studies indicate that oxytocin is relevant only in dyadic romantic relationships as opposed to group contexts (Pearce et al. 2017). OE reinforcement really makes sense only in a group context, though research that pursues a role for oxytocin in acoustic agreement would certainly be of interest. Harvey's point about dopamine is well taken. In fact, dopamine is almost always released at the same time as endorphin, and they act in something of a push-pull fashion (George and van Loon 1982); given the complexity of circumstances in which humans vocalize and their vocal production of all kinds is perceived, we remain open to new data about other neurohormones that might be involved in vocal interaction.

The commentary of Marin and Gingras deals principally with their view that OE is an exclusively sensory issue conforming to a narrow definition of auditory processing. Wagner, Bowling, and Hoeschele in their commentary similarly express the concern that our article presents "multiple definitions" of OE. It was not our intention to do so, for which clarification we acknowledge the need and will take care in future to express our conception as "manifestations of response to the perception of OE": (1) "a unique vocal property . . . whereby adult male voices are, on average, an octave lower in pitch than those of adults females and children"; (2) "a phenomenon in pitch perception [comprising] the ability to recognize that two notes are in harmony when an octave apart"; and (3) "the ability to produce octaves vocally." That said, it is difficult to imagine that one could have the one (pitch production) without the other (pitch perception; Watts, Moore, and McCaghren 2005). This does not, however, change the central claim of our article—that the capacity for OE (production *and* perception) was the product of selection pressure to enable what is in effect unison singing by adult men and women in order to bond a community.

Our article intended to consider the behavioral and productive consequences for human evolution of exploring the consequences of such a sensory response for a range of capacities dependent on the emergence of a dependable feedback loop that relates aural processing to vocalization. In more recent work (N. Bannan, R. M. Dunbar, A. R. Harvey, and P. Podlipniak, unpublished manuscript), we refine the model

presented in our current article to illustrate a Baldwinian mechanism whereby the ancient mammalian foundations of harmonic series sensitivity shaped the capacities of the human vocal system for unison and chordal exchange and mutual amplification.

Wagner, Bowling, and Hoeschele take a generally positive approach to the implications of our article, in respect of which we should acknowledge the debt it owes to their own pioneering work in this field, both collectively and individually, without which our hypothesis could not be presented. We warm to the additional information they offer that extends the clarity of our proposal, for example, “perhaps we perceive two harmonic sounds separated by an octave as the ‘same’ because they are as similar as two harmonic sounds with different fundamental frequencies can possibly be, with overlap in acoustic frequency content at a maximum.” Their proposal for further cross-species research parallels Taylor’s concerns, and we recognize the potential contribution that “an evolved sensitivity to the harmonic content of conspecific vocalizations” may have made prior to the emergence of human dimorphism. Indeed, we would be keen to consider “the (presumably) adaptive loss of secondary vocal membranes in the human larynx” in relation to the time line we propose for biological developments that may have been shaped by OE sensitivity.

Taylor’s concern regarding what she sees as an exceptionalist stance is fair comment. We fully acknowledge the extraordinary capacities of widespread species, not least that to which she has devoted much of her own research, the Australian pied butcher bird, which possesses musical abilities of astonishing clarity and variety. Indeed, we remain open to considering the important role of interaction with animals in the development of human abilities. But on lines defined by Hockett (1960), we based our approach on explaining the sequence by which specifically human vocal abilities, including dimorphism of range, may have occurred within a spectrum of “design features.” The human range *is* exceptional by comparison with that observed in any other primate. For that reason alone, it requires explanation—not because humans are somehow exceptional (every species is, after all, exceptional!) but because they stand out as different in this one particular detail. One should not confuse general explanations with what happens in specific species, because many factors affect particular biological traits and every species is going to present a particular combination of these. The question is not whether other animals can entrain or recognize pitch or even in some cases sing in unison (when the sexes are the same size) but whether any animals’ voices are pitched such that if sexes that differ in body mass sing together, they will be effortlessly in tune.

Her other claim, that there are tribes that do not sing at the octave, really depends on (1) how many do not versus do (the work of Savage and colleagues is addressing just such features); (2) whether it is really the case that central Australian natives only ever sing in 1 : 1 unison (and if so, how is this achieved—do the men remain high or women sing low? And how and why did they come to this negotiated compromise?); and (3) recognizing

that OE evolved in hominins so they could sing in “unison” does not mean some of them could not, perhaps later, have also figured out how to sing polyphony. While we agree that our article set out to explore a suite of abilities that would appear to be exceptional in comparison with other species, we were careful not to argue for intraspecies exceptionalism as others have done historically when discussing human musicality.

Indeed, the focus on OE in this article is only part of a more extensive picture that relates to our model in terms of the wider influence of harmonic series responses. Taylor is right to point out that in many human cultures other intervals determine the nature of parallel part-singing. As with the subjective and varied responses to taste that define local differences in cuisine, responses to intervallic properties in vocal repertoires represent exploitation of different valences of the same acoustic arrangement. The harmonic series presents perceptual categories that are ancient within the mammalian line but to which humans respond instinctively through the capacity to blend pitches, as well as assigning categorical linguistic roles to the differences in timbre that source filter vocal production enables.

This is again where we part company with Marin and Gingras: the “active-passive” features of the relationship between auditory perception and vocal production are not and cannot be equal, for all that the feedback loop between them defines the potential for vocal learning. While hearing sounds arises according to the precise sensory processing Marin and Gingras draw upon, human vocal production is, less predictably, both motivated and inhibited by such a wide range of independent factors (e.g., health, age, climate, emotional state, intention, etc.—all related to the complex endocrine interactions previously discussed). The collision between what we set out to explain as the real-life complexity of human vocal production and the narrow perceptual definition of OE advocated by Marin and Gingras represents a messy relationship that research in this field is obliged to negotiate. Nevertheless, we felt that this was the link that demanded to be made for a possible pathway to be proposed toward a more fine-grained model of vocal evolution.

The substantive issue here is the contrast between explanations about mechanisms and explanations about (evolutionary) function (or *how* and *why* questions) that lie at the root of all evolutionary analyses. Forming part of Tinbergen’s (1963) “Four Whys” (but dating back to Aristotle ca. 350 BCE), they have been a bulwark against confounding explanations of different logical (and biological) status. They are not alternatives but complements. All complete explanations require both. However, it is a central feature of this construction that all four questions (mechanism, function, ontogeny, and phylogeny) are independent (i.e., the answer to one does not influence the answer to any of the others) *and* we can investigate them independently without prejudice (Dunbar 2019).

Similar considerations arise in the commentary of Savage and Ozaki. We recognize the need for further detail in respect of differences in range between singing and speaking voices, and the cultural and motivational factors involved in making

comparisons, and we gratefully acknowledge the contribution made by the chart they provide. Further examples on very similar lines are in fact included in appendixes S1–S6 (available online) for our original article and, taken together with theirs, will present a convincing basis for further work. Ratios of pitch differences are unlikely ever to be exact, but that is not quite the point. We would question whether a ratio of 0.55 compared to 0.50 would be noticeably different to anyone but a trained musician, especially given the vagaries of the real-life environment where community singing (as opposed to formal concerts) takes place. More importantly, perhaps, a ratio of 0.55 is still *much* lower than you would expect for body size.

—Nicholas Bannan, Robin Dunbar, and Joshua Bamford

References Cited

- Allen, E. J., J. Mesik, K. N. Kay, and A. J. Oxenham. 2022. Distinct representations of tonotopy and pitch in human auditory cortex. *Journal of Neuroscience* 42:416–434. [ARH]
- Apicella, Coren L., David R. Feinberg, and Frank W. Marlowe. 2007. Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters* 3(6):682–684.
- Aristotle. 1927. *Prob. Xix 39*. In *The works of Aristotle*, vol. VII, *Problemata*. E. S. Forster, trans. Pp. 918b. Oxford: Clarendon.
- Arom, Simha. 2010. Corroborating external observation by cognitive data in the description and modelling of traditional music. *Musicae Scientiae* 14(2): 295–306.
- Aung, Toe, Alexander K. Hill, Dana Pfefferle, Edward McLester, James Fuller, Jenna M. Lawrence, Ivan Garcia-Nisa, et al. 2023. Group size and mating system predict sex differences in vocal fundamental frequency in anthropoid primates. *Nature Communications* 14:4069.
- Averill, Gage. 1999. Bell tones and ringing chords: sense and sensation in barbershop harmony. *World of Music* 41(1):37–51.
- Baimel, Adam, Rachel L. Severson, Andrew S. Baron, and Susan A. J. Birch. 2015. Enhancing “theory of mind” through behavioral synchrony. *Frontiers in Psychology* 6:870.
- Baldwin, J. M. 1896. A new factor in evolution. *American Naturalist* 30:441–451, 536–553.
- Bannan, Nicholas. 2003. Reverse-engineering the human voice: examining the adaptive prerequisites for song and language. In *Proceedings of the Fifth Triennial Conference of the European Society for the Cognitive Sciences of Music*. Reinhard Kopiez, Andreas C. Lehmann, Irving Wolther, and Christian Wolf, eds. P. 230. Hanover: Hochschule für Musik und Theater.
- . 2008. Language out of music: the four dimensions of vocal learning. *Australian Journal of Anthropology* 19(3):272–293.
- . 2020. An evolutionary perspective on the human capacity for singing. In *Development*, vol. 1 of *The Routledge companion to interdisciplinary studies in singing*. Frank Russo, Beatriz Ilari, and Annabel Cohen, eds. Pp. 39–51. New York: Routledge.
- Bekerman, I., P. Gottlieb, and P. Vaiman. 2014. Variations in eyeball diameters of the healthy adults. *Journal of Ophthalmology* 2014:503645. [ARH]
- Belyk, Michel, and Steven Brown. 2016. Pitch underlies activation of the vocal system during affective vocalization. *Social Cognitive and Affective Neuroscience* 11(7):1078–1088.
- Berg, M., M. Fuchs, K. Wirkner, M. Loeffler, C. Engel, and T. Berger. 2017. The speaking voice in the general population: normative data and associations to sociodemographic and lifestyle factors. *Journal of Voice* 31(2):257.e13–257.e24. <https://doi.org/10.1016/j.jvoice.2016.06.001>. [PES/YO]
- Berlioz, Hector. 1844. *Grande traité d'instrumentation et d'orchestration modernes*. Paris: Schonenberger.
- Bertolo, Mila, Daniel Müllensiefen, Isabelle Peretz, Sarah C. Woolley, Jon T. Sakata, and Samuel A. Mehr. 2023. Sex differences in human music perception are negligible. *BioRxiv*, <https://doi.org/10.1101/2023.05.23.541970>. [PES/YO]
- Blackwell, H. Richard, and Harold Schlosberg. 1943. Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *Journal of Experimental Psychology* 33(5):407.
- Borra, Tobias, Huib Versnel, Chantal Kemner, A. John van Opstal, and Raymond van Ee. 2013. Octave effect in auditory attention. *Proceedings of the National Academy of Sciences of the USA* 110(38):15225–15230.
- Bowling, D., M. Garcia, J. C. Dunn, R. Rupprecht, A. Stewart, K.-H. Frommolt, and W. Tecumseh Fitch. 2017a. Body size and vocalization in primates and carnivores. *Scientific Reports* 7(1):1–11.
- Bowling, D., M. Hoeschele, K. Z. Gill, and W. Tecumseh Fitch. 2017b. The nature and nurture of musical consonance. *Music Perception* 35(1):118–121.
- Bowling, D., and D. Purves. 2015. A biological rationale for musical consonance. *Proceedings of the National Academy of Sciences of the USA* 112(36): 11155–11160.
- Bowling, D., D. Purves, and K. Gill. 2018. Vocal similarity predicts the relative attraction of musical chords. *Proceedings of the National Academy of Sciences of the USA* 115(1):216–221.
- Brown, S. 2000. The “musilanguage” model of human evolution. In *The origins of music*. S. Brown, B. Merker, and N. L. Wallin, eds. Pp. 271–300. Cambridge, MA: MIT Press. [ARH]
- Brown, Steven, and Joseph Jordania. 2013. Universals in the world’s musics. *Psychology of Music* 41(2):229–248. <https://doi.org/10.1177/0305735611425896>. [PES/YO]
- Burns, Edward M. 1999. Intervals, scales, and tuning. In *Psychology of music*. 2nd edition. Diana Deutsch, ed. Pp. 215–264. San Diego, CA: Academic Press. [MMM/BG, HT]
- Castellengo, Michèle, Bernard Lortat-Jacob, and Gilles Léothaud. 2001. Pitch perception: five voices with four Sardinian singers. *International Symposium on Musical Acoustics*, pp. 351–355. https://www.uni-potsdam.de/fileadmin/projects/soundscapelab/DocumentsGeneral/Michèle_Bernard_Gilles_2001_PITCH_PERCEPTION_FIVE_VOICES_WITH_FOUR_SARDINIAN_SINGERS.pdf.
- Charlton, B. D., and D. Reby. 2016. The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications* 7(1):12739.
- Chiba, Gakuto, Meng-Jou Ho, Shoichiro Sato, Jiei Kuroyanagi, Joren Six, Peter Pfordresher, Adam Tierney, Shinya Fujii, and Patrick E. Savage. 2019. Small-integer ratios predominate throughout the world’s musical scales. *Southeast Asia* 2(1):1.
- Cohen, E., R. Ejsmond-Frey, N. Knight, and R. I. M. Dunbar. 2010. Rowers’ high: behavioural synchrony is correlated with elevated pain thresholds. *Biology Letters* 6:106–108.
- Cooksey, John M. 1997. Voice transformation in male adolescents. In *Bodymind and voice: foundations of voice education*. John Cooksey, Leon Thurman, and Graham Welch, eds. Pp. 718–738. Collegeville, MN: Voice Care Network.
- Cooper, N. P. 1998. Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea. *Journal of Physiology* 509:277–288. [ARH]
- Cooper, Paul. 1973. *Perspectives in music theory: an historical-analytical approach*. New York: Harper & Row. [BW/DLB/MH]
- Covey-Crump, Roger. 1992. Vocal consort style and tunings. *Companion to Contemporary Musical Thought* 2:1020–1050.
- Crook, J. H. 1972. Sexual selection, dimorphism, and social organization in the primates. In *Sexual selection and the descent of man*. Bernard Campbell, ed. Pp. 231–281. Chicago: Aldine.
- Cross, I. 2001. Music, mind and evolution. *Psychology of Music* 29:95–102. [ARH]
- . 2003. Music, cognition, culture and evolution. In *The cognitive neuroscience of music*. Isabelle Peretz and Robert J. Zatorre, eds. Pp. 42–56. Oxford: Oxford University Press. [ARH]
- Dargie, Dave. 2018. A report on the Dave Dargie collection in the International Library of African Music. <https://tranquanghaisworldthroatsinging.com/category/dave-dargie/>.
- Darwin, C. 1871. *The descent of man and selection in relation to sex*. London: J. Murray.
- Dávid-Barrett, T., and R. I. M. Dunbar. 2013. Processing power limits social group size: computational evidence for the cognitive costs of sociality. *Proceedings of the Royal Society B* 280:20131151.
- Davies, Stephan, and Matteo Ravasio. 2021. Musical universals. In *The science and psychology of music: from Beethoven at the office to Beyoncé at the gym*. William Forde Thompson and Kirk N. Olsen, eds. Pp. 47–52. Santa Barbara, CA: Greenwood. [MMM/BG]
- Della Bella, S., M. Berkowska, and J. Sowiński. 2015. Moving to the beat and singing are linked in humans. *Frontiers in Human Neuroscience* 9:663. [ARH]

- Demany, Laurent, and Françoise Armand. 1984. The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America* 83:687–695. [MMM/BG]
- Demany, L., G. Monteiro, C. Semal, S. Shamma, and R. P. Carlyon. 2021. The perception of octave pitch affinity and harmonic fusion have a common origin. *Hearing Research* 404:108213. [ARH]
- de Muinck Keizer-Schrama, S. M. P. F., and D. Mul. 2001. Trends in pubertal development in Europe. *Human Reproduction Update* 7(3):287–291.
- Deutsch, Diana, and Richard C. Boulanger. 1984. Octave equivalence and the immediate recall of pitch sequences. *Music Perception* 2(1):40–51.
- de Waal, Frans. 2016. *Are we smart enough to know how smart animals are?* London: Granta. [HT]
- Dissanayake, Ellen. 2008. If music is the food of love, what about survival and reproductive success? *Musicae Scientiae* 12(suppl.):169–195.
- Doğantan-Dack, M. 2013. Tonality: the shape of affect. *Empirical Musicology Review* 8(3–4):208–218.
- Dougherty, R. F., V. M. Koch, A. A. Brewer, B. Fischer, J. Modersitzki, and B. A. Wandell. 2003. Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *Journal of Vision* 3:586–598. [ARH]
- Druskin, Mikhail. 1983. *Igor Stravinsky: his personality, works and views*. Martin Cooper, trans. Cambridge: Cambridge University Press.
- Dunbar, R. 1996. *Grooming, gossip and the evolution of language*. Cambridge, MA: Harvard University Press. [ARH]
- Dunbar, Robin I. M. 1988. *Primate social systems (studies in behavioural adaptation)*. New York: Springer.
- . 1998. *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.
- . 2003. The social brain: mind, language, and society in evolutionary perspective. *Annual Review of Anthropology* 32(1):163–181.
- . 2009. Why only humans have language. In *The prehistory of language*. Rudi Botha and Chris Knight, eds. Pp. 12–35. Oxford: Oxford University Press.
- . 2012a. Bridging the bonding gap: the transition from primates to humans. *Philosophical Transactions of the Royal Society B* 367:1837–1846.
- . 2012b. On the evolutionary function of song and dance. In *Music, language, and human evolution*. Nicholas Bannan, ed. Pp. 201–14. Oxford: Oxford University Press. [PES/YO]
- . 2014. Mind the gap: or why humans aren't just great apes. In *Lucy to language: the benchmark papers*. R. I. M. Dunbar, C. Gamble, and J. A. J. Gowlett, eds. Pp. 3–18. Oxford: Oxford University Press.
- . 2016. *Human evolution*. New York: Oxford University Press.
- . 2019. *Evolution: what everyone needs to know*. Oxford: Oxford University Press.
- . 2020. Structure and function in human and primate social networks: implications for diffusion, network stability and health. *Proceedings of the Royal Society A* 476:20200446.
- . 2021. Virtual touch and the human social world. *Current Opinion in Behavioral Science* 43:14–19.
- . 2022. Laughter and its role in the evolution of human social bonding. *Philosophical Transactions of the Royal Society B* 377:20210176.
- Dunbar, Robin I. M., Kostas Kaskatis, Ian MacDonald, and Vinnie Barra. 2012. Performance of music elevates pain threshold and positive affect: implications for the evolutionary function of music. *Evolutionary Psychology* 10(4):147470491201000403.
- Dunbar, R. I. M., E. Pearce, B. Tarr, A. Makdani, J. Bamford, S. Smith, and F. McGlone. 2021. Cochlear SGN neurons elevate pain thresholds in response to music. *Scientific Reports* 11:14547.
- Dunbar, Robin I. M., and Susanne Shultz. 2021a. The infertility trap: the fertility costs of group-living in mammalian social evolution. *Frontiers in Ecology and Evolution* 9:634664.
- . 2021b. Social complexity and the fractal structure of social groups in primate social evolution. *Biology Reviews* 96:1889–1906.
- Dunn, Jacob C., Lauren B. Halenar, Thomas G. Davies, Jurgi Cristobal-Azkarate, David Reby, Dan Sykes, Sabine Dengg, W. Tecumseh Fitch, and Leslie A. Knapp. 2015. Evolutionary trade-off between vocal tract and testes dimensions in howler monkeys. *Current Biology* 25:2839–2844.
- Echternach, Matthias, Michael Mark, and Bernhard Richter. 2011. Vocal tract configurations in yodelling—prospective comparison of two Swiss yodeller and two non-yodeller subjects. *Logopedics Phoniatrics Vocology* 36(3):109–113.
- Falk, Dean. 2004. Prelinguistic evolution in early hominins: whence motherese? *Behavioral and Brain Sciences* 27(4):491–583.
- Falkenhausen, Lothar von. 1992. On the early development of Chinese musical theory: the rise of pitch-standards. *Journal of the American Oriental Society* 112:433–439.
- Fancourt, Daisy, Lisa Aufegger, and Aaron Williamson. 2015. Low-stress and high-stress singing have contrasting effects on glucocorticoid response. *Frontiers in Psychology* 6:1242.
- Faulkner, Robert. 2012. Icelandic men, male voice choirs and masculine identity. In *Perspectives on males and singing*. Scott Harrison, Graham Welch, and Adam Adler, eds. Pp. 215–232. Dordrecht: Springer.
- Fischer, Julia, Kurt Hammerschmidt, Dorothy L. Cheney, and Robert M. Seyfarth. 2002. Acoustic features of male baboon loud calls: influences of context, age, and individuality. *Journal of the Acoustic Society of America* 111:1465–1474.
- Fitch, W. Tecumseh. 2002. Comparative vocal production and the evolution of speech: reinterpreting the descent of the larynx. In *The transition to language*. Alison Wray, ed. Pp. 21–45. Oxford: Oxford University Press.
- . 2006. The biology and evolution of music: a comparative perspective. *Cognition* 100(1):173–215. [PES/YO]
- Fitch, W. Tecumseh, and Mark Hauser. 2003. Unpacking “honesty”: vertebrate vocal production and the evolution of acoustic signals. In *Acoustic communication*. Arthur Popper, Richard R. Fay, and Andrea Megela Simmons, eds. Pp. 65–137. New York: Springer.
- Fitch, W. Tecumseh, and Jay Giedd. 1999. Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *Journal of the Acoustical Society of America* 106(3):1511–1522.
- Fitch, W. Tecumseh, and David Reby. 2001. The descended larynx is not uniquely human. *Proceedings of the Royal Society B* 268(1477):1669–1675.
- Foss, Alexander, Eric Altschuler, and Karin James. 2007. Neural correlates of the Pythagorean ratio rules. *NeuroReport* 18(15):1521–1525.
- Fujiwara, T., O. Weisman, M. Ochi, K. Shirai, K. Matsumoto, E. Noguchi, and R. Feldman. 2019. Genetic and peripheral markers of the oxytocin system and parental care jointly support the cross-generational transmission of bonding across three generations. *Psychoneuroendocrinology* 2:172–181. [ARH]
- Gabisiona, T. 2015. Semiotic dimensions of drone in traditional music. In *The Seventh International Symposium on Traditional Polyphony, proceedings*. Pp. 141–147. Georgia: International Research Center for Traditional Polyphony.
- Gackle, Lynne. 1991. The adolescent female voice characteristics of change and stages of development. *Choral Journal* 31(8):17–25.
- Garnett, Liz. 1999. Separate but equal? sexual politics in the barbershop. *Women and Music* 3:28–42.
- Gaudzinski-Windheuser, S., E. S. Noack, E. Pop, C. Herbst, J. Pflöging, J. Buchli, et al. 2018. Evidence for close-range hunting by last interglacial Neanderthals. *Nature Ecology and Evolution* 2(7):1087–1092.
- Gaulin, Steven, and James Boster. 1985. Cross-cultural differences in sexual dimorphism: is there any variance to be explained? *Ethology and Sociobiology* 6(4):219–225.
- George, S. R., and G. R. van Loon. 1982. β -endorphin alters dopamine uptake by the dopamine neurons of the hypothalamus and striatum. *Brain Research* 248(2):293–303.
- Ghazanfar, Asif A., and Drew Rendall. 2008. Evolution of human vocal production. *Current Biology* 18(11):R457–R460.
- Gill, Kamraan Z., and Dale Purves. 2009. A biological rationale for musical scales. *PLoS ONE* 4(12):e8144. [MMM/BG]
- Grape, C., M. Sandgren, L. O. Hansson, M. Ericson, and T. Theorell. 2003. Does singing promote well-being? an empirical study of professional and amateur singers during a singing lesson. *Integrative Physiology and Behavioral Science* 38:65–74. [ARH]
- Gruss, L. T., and D. Schmitt. 2015. The evolution of the human pelvis: changing adaptations to bipedalism, obstetrics and thermoregulation. *Philosophical Transactions of the Royal Society B* 370(1663):20140063.
- Gursul, D., S. Goksan, C. Hartley, G. S. Mellado, F. Moultrie, A. Hoskin, E. Adams, et al. 2018. Stroking modulates noxious-evoked brain activity in human infants. *Current Biology* 28(24):R1380–R1381.
- Hagel, S. 2009. *Ancient Greek music: a new technical history*. Cambridge: Cambridge University Press.
- Hamilton, William D. 1971. Geometry for the selfish herd. *Journal of Theoretical Biology* 31(2):295–311.
- Harcourt, Alexander H., Paul H. Harvey, Susan G. Larson, and Roger V. Short. 1981. Testis weight, body weight and breeding system in primates. *Nature* 293(5827):55–57.
- Harries, Meredydd L., Judith M. Walker, David M. Williams, S. Hawkins, and I. A. Hughes. 1997. Changes in the male voice at puberty. *Archives of Disease in Childhood* 77(5):445–447.

- Harvey, Alan R. 2020. Links between the neurobiology of oxytocin and human musicality. *Frontiers in Human Neuroscience* 14:350.
- . 2017. *Music, evolution, and the harmony of souls*. Oxford: Oxford University Press.
- He, W., and T. Ren. 2021. The origin of mechanical harmonic distortion within the organ of Corti in living gerbil cochleae. *Communications Biology* 4:1008. [ARH]
- Helmholtz, Hermann von. 1862. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: Vieweg. [MMM/BG]
- Henrich, Nathalie. 2006. Mirroring the voice from Garcia to the present day: some insights into singing voice registers. *Logopedics Phoniatrics Vocology* 31(1):3–14.
- Hilton, Courtney B., Cody J. Moser, Mila Bertolo, Harry Lee-Rubin, Dorsa Amir, Constance M. Bainbridge, Jan Simson, et al. 2022. Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour* 6:1545–1556. <https://doi.org/10.1038/s41562-022-01410-x>. [PES/YO]
- Hockett, Charles. 1960. The origin of speech. *Scientific American* 203(3):88–97.
- Hoeschele, Marisa. 2017. Animal pitch perception: melodies and harmonies. *Comparative Cognition and Behavior Reviews* 12:5–18.
- Hoeschele, Marisa, Ronald G. Weisman, Lauren M. Guillette, Allison H. Hahn, and Christopher B. Sturdy. 2013. Chickadees fail standardized operant tests for octave equivalence. *Animal Cognition* 16(4):599–609.
- Hoeschele, Marisa, Ronald G. Weisman, and Christopher B. Sturdy. 2012. Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. *Attention, Perception, and Psychophysics* 74(8):1742–1760.
- Honing, Henkjan, Carel ten Cate, Isabelle Peretz, and Sandra E. Trehub. 2015. Without it no music: cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society B* 370:20140088.
- Hughes, Susan M., Franco Dispenza, and Gordon G. Gallup Jr. 2004. Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior* 25(5):295–304.
- Hughes, Susan M., Sally D. Farley, and Bradley C. Rhodes. 2010. Vocal and physiological changes in response to the physical attractiveness of conversational partners. *Journal of Nonverbal Behavior* 34(3):155–167.
- Hyde, K. L., J. P. Lerch, R. J. Zatorre, T. D. Griffiths, A. C. Evans, and I. Peretz. 2007. Cortical thickness in congenital amusia: when less is better than more. *Journal of Neuroscience* 21:13028–13032. [ARH]
- Hyde, K. L., R. J. Zatorre, T. D. Griffiths, J. P. Lerch, and I. Peretz. 2006. Morphometry of the amusic brain: a two-site study. *Brain* 129:2562–2570. [ARH]
- Ikhtisamov, K. S. 1988. On the problem of comparative investigation of two-part throat singing and instrumental music of Turkic and Mongol peoples. In *Folk musical instruments and instrumental music*. Y. Gippius, ed. Pp. 197–216. Moscow: Sovetskii Kompozitor.
- Jaatinen, J., V. Vantänen, V. Salmela, and K. Alho. 2023. Subjectively preferred octave size is resolved at the late stages of cerebral auditory processing. *European Journal of Neuroscience* 58:3686–3704. [ARH]
- Jacoby, Nori, Eduardo A. Undurraga, Malinda J. McPherson, Joaquín Valdés, Tomás Ossandón, and Josh H. McDermott. 2019. Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology* 29(19):3229–3243.
- Jones, J. A., and D. Keough. 2008. Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research* 190:279–287.
- Jordania, Joseph. 2006. *Who asked the first question? the origins of human choral singing, intelligence, language and speech*. Tbilisi: Logos.
- Keeler, J. R., E. A. Roth, B. L. Neuser, J. M. Spitsbergen, D. J. Waters, and J. M. Vianney. 2015. The neurochemistry and social flow of singing: bonding and oxytocin. *Frontiers in Human Neuroscience* 9:518. [ARH]
- Keller, Peter E., Rasmus König, and Giacomo Novembre. 2017. Simultaneous cooperation and competition in the evolution of musical behavior: sex-related modulations of the singer's formant in human chorusing. *Frontiers in Psychology* 8(1559):1–14.
- Kim, S., P. Fonagy, O. Koos, K. Dorsett, and L. Strathearn. 2014. Maternal oxytocin response predicts mother-to-infant gaze. *Brain Research* 1580:133–142. [ARH]
- Kingma, Sjouke A., Peter Santema, Michael Taborsky, and Jan Komdeur. 2014. Group augmentation and the evolution of cooperation. *Trends in Ecology and Evolution* 29:476–484.
- Knight, Chris, and Jerome Lewis. 2017. Wild voices: mimicry, reversal, metaphor, and the emergence of language. *Current Anthropology* 58(4):435–453.
- Knowles, K. K., and A. C. Little. 2016. Vocal fundamental and formant frequencies affect perceptions of speaker cooperativeness. *Quarterly Journal of Experimental Psychology* 69(9):1657–1675. <https://doi.org/10.1080/17470218.2015.1091484>. [PES/YO]
- Kokko, Hanna, Rufus A. Johnstone, and Tim Clutton-Brock. 2001. The evolution of cooperative breeding through group augmentation. *Proceedings of the Royal Society B* 268(1463):187–196.
- Korner, Anneliese, and Evelyn Thoman. 1972. The relative efficacy of contact and vestibular-proprioceptive stimulation in soothing neonates. *Child Development* 43(2):443–453.
- Kreutz, G. 2014. Singing and social bonding. *Music and Medicine* 6:51–60. [ARH]
- Kumar, V., P. L. Croxson, and K. Simonyan. 2016. Structural organization of the laryngeal motor cortical network and its implication for evolution of speech production. *Journal of Neuroscience* 36:4170–4418. [ARH]
- Kuroyanagi, Jiei, Shoichiro Sato, Meng-Jou Ho, Gakuto Chiba, Joren Six, Peter Pfordresher, Adam Tierney, Shinya Fujii, and Patrick Savage. 2019. Automatic comparison of human music, speech, and bird song suggests uniqueness of human scales. *9th Folk Music Analysis Conference*. PsyArXiv, <https://doi.org/10.31234/osf.io/zpv5w>.
- Larsson, Matz, Joachim Richter, and Andrea Ravignani. 2019. Bipedal steps in the development of rhythmic behavior in humans. *Music and Science* 2:2059204319892617.
- Lehmann, Arthur C., and Louis J. Mihalyi. 1982. Aggression, bravery, endurance, and drugs: a radical re-evaluation and analysis of the Masai warrior complex. *Ethnology* 21(4):335–347.
- Lehmann, Julia, Amanda H. Korstjens, and Robin I. M. Dunbar. 2007. Group size, grooming and social cohesion in primates. *Animal Behaviour* 74(6):1617–1629.
- Lewis, Z., and P. J. Sullivan. 2018. The effect of group size and synchrony on pain threshold changes. *Small Group Research* 49(6):723–738.
- List, George. 1984. Concerning the concept of the universal and music. *World of Music* 26(2):40–49.
- Livingstone, F. B. 1973. Did the Australopithecines sing? *Current Anthropology* 14(1/2):25–29.
- Livingstone, Steven Robert, and William Forde Thompson. 2009. The emergence of music from the theory of mind. *Musicae Scientiae* 13(suppl.):83–115.
- Luo, Zhe-Xi, and Geoffrey A. Manley. 2020. Origins and early evolution of mammalian ears and hearing function. In *The senses: a comprehensive reference*, vol. 2. 2nd edition. Benedikt Grothe and Bernd Fritzsche, eds. Pp. 207–252. Elsevier: Amsterdam. [MMM/BG]
- Lyon, Bruce E., and Robert Montgomerie. 2012. Sexual selection is a form of social selection. *Philosophical Transactions of the Royal Society B* 367(1600):2266–2273.
- Mandell, J., K. Schulze, and G. Schlaug. 2007. Congenital amusia: an auditory-motor feedback disorder? *Restorative Neurology and Neuroscience* 25:323–334. [ARH]
- Manninen, Sandra, Lauri Tuominen, Robin I. Dunbar, Tomi Karjalainen, Jussi Hirvonen, Eveliina Arponen, Riitta Hari, Iiro P. Jääskeläinen, Mikko Sams, and Lauri Nummenmaa. 2017. Social laughter triggers endogenous opioid release in humans. *Journal of Neuroscience* 37(25):6125–6131.
- Martin, Robert D. 1990. *Primate origins and evolution*. London: Chapman & Hall.
- McDermott, Josh H., Alan F. Schultz, Eduardo A. Undurraga, and Ricardo A. Godoy. 2016. Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* 535(7613):547–550.
- McPherson, T., D. Berger, S. Alagapan, and F. Frölich. 2018. Intrinsic rhythmicity predicts synchronization-continuation entrainment performance. *Scientific Reports* 8:11782. [ARH]
- McPherson, M. J., S. E. Dolan, A. Durango, T. Ossandón, J. Valdés, E. A. Undurraga, N. Jacoby, R. A. Godoy, and J. H. McDermott. 2020. Perceptual fusion of musical notes by native Amazonians suggests universal representations of musical intervals. *Nature Communications* 11:2786. [ARH]
- Mehr, Samuel A., Max M. Krasnow, Gregory A. Bryant, and Edward H. Hagen. 2020. Origins of music in credible signaling. *Behavioral and Brain Sciences* 44:e60.
- . 2021. Origins of music in credible signaling. *Behavioral and Brain Sciences* 44(e59):23–39. <https://doi.org/10.1017/S0140525X20000345>. [PES/YO]
- Mehr, Samuel A., Manvir Singh, Dean Knox, Daniel M. Ketter, Daniel Pickens-Jones, Stephanie Atwood, Christopher Lucas, et al. 2019. Univer-sality and diversity in human song. *Science* 366:eaax0868. <https://doi.org/10.1126/science.aax0868>. [PES/YO]

- Merker, Bjorn. 2000. Synchronous chorusing and human origins. In *The origins of music*. Nils Wallin, Bjorn Merker, and Steven Brown, eds. Pp. 315–327. Cambridge, MA: MIT Press.
- . 2012. The vocal learning constellation: Imitation, ritual culture, encephalization. In *Music, language, and human evolution*. Nicholas Bannan, ed. Pp. 215–60. Oxford: Oxford University Press.
- . 2021. Music, bonding, and human evolution: a critique. *Behavioural Brain Science* 44:e83. [ARH]
- Mersenne, Marin. 1636/1975. *Harmonie universelle: contenant la théorie et la pratique de la musique*. Paris: Editions du Centre National de la Recherche Scientifique.
- Miller, Donald Gray. 2000. *Registers in singing: empirical and systematic studies in the theory of the singing voice*. Groningen: Rijksuniversiteit.
- Miller, Geoffrey. 2000a. *The mating mind: how sexual choice shaped the evolution of human nature*. London: Heinemann.
- . 2000b. Evolution of human music through sexual selection. In *The origins of music*. Nils L. Wallin, Bjorn Merker, and Steven Brown, eds. Pp. 329–60. Cambridge, MA: MIT Press. [PES/YO]
- Mithen, Steven. 2005. *The singing Neanderthals: the origins of music, language, mind and body*. London: Weidenfeld & Nicolson.
- Moreau, P., P. Jolicœur, and I. Peretz. 2013. Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials. *Brain and Cognition* 81:337–344. [ARH]
- Morley, Iain. 2017. Pretend play, cognition, and life history in human evolution. In *Ritual, play, and belief in evolution and early human societies*. Colin Renfrew, Iain Morley, and Michael Boyd, eds. Pp. 66–86. Cambridge: Cambridge University Press.
- . 2013. *The prehistory of music: human evolution, archaeology, and the origins of musicality*. Oxford: Oxford University Press.
- Nelson, E., C. Rolian, L. Cashmore, and S. Shultz. 2011. Digit ratios predict polygyny in early apes, *Ardipithecus*, Neanderthals and early modern humans but not in *Australopithecus*. *Proceedings of the Royal Society B* 278:1556–1563.
- Nikolsky, A. 2015. Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. 1. Prehistoric. *Frontiers in Psychology* 6:1405.
- Nishimura, T., I. T. Tokuda, S. Miyachi, J. C. Dunn, C. T. Herbst, K. Ishimura, A. Kaneko, et al. 2022. Evolutionary loss of complexity in human vocal anatomy as an adaptation for speech. *Science* 377(6607):760–763, <https://doi.org/10.1126/science.abm1574> [BW/DLB/MH]
- O'Connor, J. J. M., and P. Barclay. 2017. The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior* 38(4):506–512, <https://doi.org/10.1016/j.evolhumbehav.2017.03.001> [PES/YO]
- Okasha, Samir. 2006. *Evolution and the levels of selection*. Oxford: Oxford University Press.
- Ooishi, Y., H. Mukai, K. Watanabe, S. Kawato, and M. Kashino. 2017. Increase in salivary oxytocin and decrease in salivary cortisol after listening to relaxing slow-tempo and exciting fast-tempo music. *PLoS ONE* 12:e0189075. [ARH]
- Ozaki, Yuto, Adam Tierney, Peter Q. Pfordresher, John McBride, Emmanouil Benetos, Polina Proutskouva, Gakuto Chiba, et al. 2022. Globally, songs and instrumental melodies are slower, higher, and use more stable pitches than speech (stage 2 registered report). *Peer Community in Registered Reports*, <https://doi.org/10.31234/osf.io/jr9x7>. [PES/YO]
- Paletta, P., N. Bass, M. Kavaliars, and E. Choleris. 2022. The role of oxytocin in shaping complex social behaviours: possible interactions with other neuromodulators. *Philosophical Transactions of the Royal Society B* 377: 20210058. [ARH]
- Pankovski, Toso, and Ana Pankovska. 2022. Nonspecific hebbian neural network model predicts musical scales discreteness and just intonation without using octave-equivalence mapping. *Scientific Reports* 12:8795. [MMM/BG]
- Parncutt, Richard. 2012. *Harmony: a psychoacoustical approach*. Berlin: Springer.
- Patel, Aniruddh D. 2008. *Music, language and the brain*. Oxford: Oxford University Press. [ARH, PES/YO]
- Pearce, Eiluned, Jacques Launay, and Robin I. M. Dunbar. 2015. The ice-breaker effect: singing mediates fast social bonding. *Royal Society Open Science* 2(10):150221.
- Pearce, Eiluned, Jacques Launay, Max van Duijn, Anna Rotkirch, Tamas David-Barrett, and Robin I. M. Dunbar. 2015. Singing together or apart: the effect of competitive and cooperative singing on social bonding within and between sub-groups of a university fraternity. *Psychology of Music* 44(6): 1255–1273.
- Pearce, Eiluned, Andrew Shuttleworth, Matt Grove, and R. H. Layton. 2014. The costs of being a high-latitude hominin. In *Lucy to language: the benchmark papers*. R. I. M. Dunbar, C. Gamble, and J. A. J. Gowlett, eds. Pp. 356–379. Oxford: Oxford University Press.
- Pearce, E., C. Stringer, and R. I. M. Dunbar. 2013. New insights into differences in brain organisation between Neanderthals and anatomically modern humans. *Proceedings of the Royal Society B* 280:1471–1481.
- Pearce, E., R. Wlodarski, A. J. Machin, and R. I. M. Dunbar. 2017. Variation in the β -endorphin, oxytocin, and dopamine receptor genes is associated with different dimensions of human sociality. *Proceedings of the National Academy of Sciences of the USA* 114(20):5300–5305, <https://doi.org/10.1073/pnas.1700712114>.
- Peter, B., B. Foster, H. Haas, K. Middleton, and K. McKibben. 2015. Direct and octave-shifted pitch matching during nonword imitations in men, women, and children. *Journal of Voice* 29:260.e21–260.e30, <https://doi.org/10.1016%2Fj.jvoice.2014.06.011>. [BW/DLB/MH]
- Peter, Beate, Tara Larkin, and Carol Stoel-Gammon. 2009. Octave-shifted pitch matching in nonword imitations: the effects of lexical stress and speech sound disorder. *Journal of the Acoustical Society of America* 126(4): 1663–1666.
- Peter, Beate, Carol Stoel-Gammon, and Daniella Kim. 2008. Octave equivalence as an aspect of stimulus-response similarity during nonword and sentence imitations in young children. *Speech Prosody* 2008:731–734.
- Pfefferle, D., and J. Fischer. 2006. Sounds and size: identification of acoustic variables that reflect body size in hamadryas baboons, *Papio hamadryas*. *Animal Behaviour* 72(1):43–51, <https://doi.org/10.1016/j.anbehav.2005.08.021>.
- Pika, Simone, Ray Wilkinson, Kobin H. Kendrick, and Sonja C. Vernes. 2018. Taking turns: bridging the gap between human and animal communication. *Proceedings of the Royal Society B* 285(1880):20180598.
- Pinker, Steven. 1984. *Language learnability and language development*. Cambridge, MA: Harvard University Press. [MMM/BG]
- . 1997. *How the mind works*. New York: Norton.
- Pisanski, Katarzyna, Paul J. Fraccaro, Cara C. Tigue, Jillian J. M. O'Connor, Susanne Röder, Paul W. Andrews, Bernhard Fink, Lisa M. DeBruine, Benedict C. Jones, and David R. Feinberg. 2014. Vocal indicators of body size in men and women: a meta-analysis. *Animal Behaviour* 95:89–99. [MMM/BG]
- Pisanski, Katarzyna, Agata Groyecka-Bernard, and Piotr Sorokowski. 2021. Human voice pitch measures are robust across a variety of speech recordings: methodological and theoretical implications. *Biology Letters* 17(9):20210356.
- Pisanski, Katarzyna, Jordan Raine, and David Reby. 2020. Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries. *Royal Society Open Science* 7(2):191642.
- Pitcairn, T. K., and M. Schleidt. 1976. Dance and decision an analysis of a courtship dance of the Medpa, New Guinea. *Behaviour* 58:298–316.
- Plavcan, J. Michael, and Carel P. Van Schaik. 1997. Interpreting hominid behavior on the basis of sexual dimorphism. *Journal of Human Evolution* 32(4):345–374.
- Podlipniak, Piotr. 2016. The evolutionary origin of pitch centre recognition. *Psychology of Music* 44(3):527–543.
- . 2017. The role of the Baldwin effect in the evolution of human musicality. *Frontiers in Neuroscience* 11:542.
- . 2020. Pitch syntax as an evolutionary prelingual innovation. *Musicae Scientiae* 26:280–302.
- Praetorius, Michael. 1958 (1618). *Syntagma musicum. II. De Organographia*. Wolfenbüttel. *Faksimile-Nachdruck*. Kassel: Bärenreiter.
- Putnam, P. T., and S. W. C. Chang. 2022. Interplay between the oxytocin and opioid systems in regulating social behaviour. *Philosophical Transactions of the Royal Society B* 377:20210050. [ARH]
- Putts, David A., Coren L. Apicella, and Rodrigo A. Cárdenas. 2012. Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B* 279(1728):601–609, <https://doi.org/10.1098/rspb.2011.0829>. [PES/YO]
- Putts, David A., Steven J. C. Gaulin, and Katherine Verdolini. 2006. Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior* 27(4):283–296.
- Putts, David A., Alexander K. Hill, Drew H. Bailey, Robert S. Walker, Drew Rendall, John R. Wheatley, Lisa L. M. Welling, et al. 2016. Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proceedings of the Royal Society B* 283(1829):20152830.
- Raine, Jordan, Katarzyna Pisanski, Rod Bond, Julia Simmer, and David Reby. 2019. Human roars communicate upper-body strength more effectively than do screams or aggressive and distressed speech. *PLoS ONE* 14(3):e0213034.

- Raine, Jordan, Katarzyna Pisanski, and David Reby. 2017. Tennis grunts communicate acoustic cues to sex and contest outcome. *Animal Behaviour* 130:47–55.
- Rameau, J.-P. 1722. *Traité de l'harmonie*. Paris: Ballard.
- Repp, B. H., and Y.-H. Su. 2013. Sensorimotor synchronization: a review of recent research (2006–2012). *Psychonomic Bulletin and Review* 20:403–452. [ARH]
- Richards, Douglas G., James P. Wolz, and Louis M. Herman. 1984. Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*. *Journal of Comparative Psychology* 98(1):10–28.
- Robledo, Juan P., Esteban Hurtado, Felipe Prado, Domingo Román, and Carlos Cornejo. 2016. Music intervals in speech: psychological disposition modulates ratio precision among interlocutors' nonlocal f0 production in real-time dyadic conversation. *Psychology of Music* 44(6):1404–1418.
- Rosenberg, Karen R. 1992. The evolution of modern human childbirth. *American Journal of Physical Anthropology* 35(S15):89–124.
- Rubenstein, Dustin. 2012. Sexual and social competition: broadening perspectives by defining female roles. *Philosophical Transactions of the Royal Society B* 367(1600):2248–2252.
- Sachs, Curt. 1943. *The rise of music in the ancient world, East and West*. New York: W. W. Norton. [HT]
- Savage, Patrick E., Steven Brown, Emi Sakai, and Thomas E. Currie. 2015. Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the USA* 112(29):8987–8992, <https://doi.org/10.1073/pnas.1414495112>. [PES/YO]
- Savage, Patrick E., Psyche Loui, Bronwyn Tarr, Adena Schachner, Luke Glowacki, Steven Mithen, and W. Tecumseh Fitch. 2021. Music as a co-evolved system for social bonding. *Behavioral and Brain Sciences* 44(e59): 1–22, <https://doi.org/10.1017/S0140525X20000333>.
- Saxton, Tamsin K., Lauren L. Mackey, Kristofor McCarty, and Nick Neave. 2016. A lover or a fighter? opposing sexual selection pressures on men's vocal pitch and facial hair. *Behavioral Ecology* 27(2):512–519.
- Schacht, Ryan, and Karem L. Kramer. 2019. Are we monogamous? a review of the evolution of pair-bonding in humans and its contemporary variation cross-culturally. *Frontiers in Ecology and Evolution* 7:230.
- Scherer, Klaus R. 1992. Vocal affect expression as symptom, symbol, and appeal. In *Nonverbal vocal communication: comparative and developmental approaches*. Hanus Papoušek, Uwe Jürgens, and Mechtild Papoušek, eds. Pp. 43–60. Paris/Cambridge: Editions de la Maison des Sciences de l'Homme/Cambridge University Press.
- Schild, Christoph, Julia Stern, and Ingo Zettler. 2020. Linking men's voice pitch to actual and perceived trustworthiness across domains. *Behavioral Ecology* 31(1):164–175, <https://doi.org/10.1093/beheco/arz173>.
- Schwartz, David A., Catherine Q. Howe, and Dale Purves. 2003. The statistical structure of human speech sounds predicts musical universals. *Journal of Neuroscience* 23(18):7160–7168.
- Schwarzkopf, D. S., C. Song, and G. Rees. 2011. The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience* 14:28–30. [ARH]
- Seltzer, L. J., T. E. Ziegler, and S. D. Pollak. 2010. Social vocalizations can release oxytocin in humans. *Proceedings of the Royal Society B* 277:2661–2666. [ARH]
- Sergeant, Desmond C., Peta J. Sjölander, and Graham F. Welch. 2005. Listeners' identification of gender differences in children's singing. *Research Studies in Music Education* 24(1):28–39.
- Sergeant, Desmond C., and Graham Welch. 2009. Gender differences in long-term average spectra of children's singing voices. *Journal of Voice* 23(3):319–336.
- Sethares, William A. 1993. Local consonance and the relationship between timbre and scale. *Journal of the Acoustical Society of America* 94(3):1218–1226.
- Shepard, Roger N. 1982. Geometrical approximations to the structure of musical pitch. *Psychological Review* 89(4):305–333. [MMM/BG]
- Shofner, W. P. 2023. Cochlear tuning and the peripheral representation of harmonic sounds in mammals. *Journal of Comparative Physiology A* 209:145–161. [ARH]
- Shultz, S., and R. I. Dunbar. 2022. Socioecological complexity in primate groups and its cognitive correlates. *Philosophical Transactions of the Royal Society B* 377(1860):20210296.
- Shultz, S., C. Opie, E. Nelson, Q. D. Atkinson, and R. Dunbar. 2014. Evolution of primate social systems: implications for hominin social evolution. In *Lucy to language: the benchmark papers*. R. I. M. Dunbar, C. Gamble, and J. A. J. Gowlett, eds. Pp. 317–322. Oxford: Oxford University Press.
- Snowdon, Charles T., Elke Zimmermann, and Eckart Altenmüller. 2015. Music evolution and neuroscience. *Progress in Brain Research* 217:17–34.
- Storr, A. 1992. *Music and the mind*. London: Harper Collins. [ARH]
- Story, Brad H., Ingo R. Titze, and Eric A. Hoffman. 2001. The relationship of vocal tract shape to three voice qualities. *Journal of the Acoustical Society of America* 109(4):1651–1667.
- Strathearn, L., P. Fonagy, J. Amico, and P. R. Montague. 2009. Adult attachment predicts maternal brain and oxytocin response to infant cues. *Neuropsychopharmacology* 34:2655–2666. [ARH]
- Stravinsky, Igor. 1962. *Autobiography*. New York: Norton.
- Sundberg, Johan. 1987. *The science of singing*. Dekalb: Northern Illinois University Press.
- Sutcliffe, Alistair, Robin Dunbar, Jens Binder, and Holly Arrow. 2012. Relationships and the social brain: integrating psychological and evolutionary perspectives. *British Journal of Psychology* 103(2):149–168.
- Tarr, B., J. Launay, E. Cohen, and R. I. M. Dunbar. 2015. Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. *Biology Letters* 11:20150767.
- Tarr, Bronwyn, Jacques Launay, and Robin I. M. Dunbar. 2016a. Music and social bonding: “self-other” merging and neurohormonal mechanisms. *Frontiers in Psychology* 5:1096.
- . 2016b. Silent disco: dancing in synchrony leads to elevated pain thresholds and social closeness. *Evolution and Human Behavior* 37:343–349.
- Thompson, W. F., M. M. Marin, and L. Stewart. 2012. Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences of the USA* 109:19027–19032. [ARH]
- Thorpe, W. H. 1973. Duet-singing birds. *Scientific American* 229:70–79.
- Tinbergen, N. 1963. On the aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20:410–433.
- Titze, Ingo R. 1994. *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.
- . 2003. The voice as a musical instrument: fundamental differences between man-made and biological designs. In *Proceedings of the Stockholm Music Acoustics Conference*. Pp. 21–24. Stockholm: KTH.
- . 2017. Human speech: a restricted use of the mammalian larynx. *Journal of Voice* 31(2):135–141.
- Tobias, Joseph A., Robert Planqué, Dominic L. Cram, and Nathalie Seddon. 2014. Species interactions and the structure of complex communication networks. *Proceedings of the National Academy of Sciences of the USA* 111:1020–1025. [HT]
- Trainor, Laurel J. 2015. The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philosophical Transactions of the Royal Society B* 370:20140089. [MMM/BG]
- Trần Quang, Hai. 2002. Polyphony in one throat. In *Proceedings of the First International Symposium on Traditional Polyphony*. Rusudan Tsursumia and Joseph Jordania, eds. Pp. 274–287. Tbilisi: International Research Center for Traditional Polyphony of Tbilisi State Conservatoire.
- Trần Quang, Hai, and Nicholas Bannan. 2012. Vocal traditions of the world: towards an evolutionary account of voice production in music. In *Music, language and human evolution*. Nicholas Bannan, ed. Pp. 142–172. Oxford: Oxford University Press.
- Trần Van, Khe. 1977. Is the pentatonic universal? a few reflections on pentatonism. *World of Music* 19(1/2):76–84.
- Trimble, Michael. 2014. *Why humans like to cry: tragedy, evolution, and the brain*. Oxford: Oxford University Press.
- van der Merwe, Peter. 1989. *Origins of the popular style: the antecedents of twentieth-century popular music*. Oxford: Clarendon.
- Van Essen, D. C., M. F. Glasser, D. I. Dierker, J. Harwell, and T. Coalson. 2012. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cerebral Cortex* 22:2241–2262. [ARH]
- Wagner, Bernhard, and Marisa Hoeschele. 2022. The links between pitch, timbre, musicality and social bonding from cross-species research. *Comparative Cognition and Behaviour Reviews* 17:13–32. [MMM/BG, BW/DLB/MH]
- Wagner, Bernhard, Dan C. Mann, Shahrzad Afroozeh, Gabriel Staubmann, and Marisa Hoeschele. 2019. Octave equivalence perception is not linked to vocal mimicry: budgerigars fail standardized operant tests for octave equivalence. *Behaviour* 156(5–8):479–504.
- Wagner, Bernhard, Vedrana Šlipogor, Jinook Oh, Marion Varga, and Marisa Hoeschele. 2023. A comparison between common marmosets (*Callithrix jacchus*) and human infants sheds light on traits proposed to be at the root of human octave equivalence. *Developmental Science* 26:e13395. [MMM/BG]
- Wagner B., C. B. Sturdy, R. G. Weisman, and M. Hoeschele. 2022. Pitch chroma information is processed in addition to pitch height information with more than two pitch-range categories. *Attention, Perception, and*

- Psychophysics* 84(5):1757–1771, <https://doi.org/10.1163/10.3758/s13414-022-02496-1>. [BW/DLB/MH]
- Walker, K. M. M., R. Gonzalez, J. Z. Kang, J. H. McDermott, and A. J. King. 2019. Across-species differences in pitch perception are consistent with differences in cochlear filtering. *eLife* 8:e1626. [ARH]
- Wallin, Nils L. 1991. *Biomusicology: neurophysiological, neuropsychological, and evolutionary perspectives on the origins and purposes of music*. Stuyvesant, NY: Pendragon.
- Warren, Jason D., Stefan Uppenkamp, Roy D. Patterson, and Timothy D. Griffiths. 2003. Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the USA* 100(17):10038–10042.
- Watts, C., R. Moore, and K. McCaghren. 2005. The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. *Journal of Voice* 19(4):534–543.
- Weinstein, Daniel, Jacques Launay, Eiluned Pearce, Robin I. M. Dunbar, and Lauren Stewart. 2016. Singing and social bonding: changes in connectivity and pain threshold as a function of group size. *Evolution and Human Behavior* 37:152–158.
- Weiss, Desco. 1950. The pubertal change of the human voice (mutation). *Folia Phoniatica et Logopaedica* 2(3):126–159.
- Welch, Graham F., and David M. Howard. 2002. Gendered voice in the cathedral choir. *Psychology of Music* 30(1):102–120.
- Weusthoff, Sarah, Brian R. Baucom, and Kurt Hahlweg. 2013. The siren song of vocal fundamental frequency for romantic relationships. *Frontiers in Psychology* 4:439.
- Will, Udo. 1997. Two types of octave relationships in central Australian vocal music? *Musicology Australia* 20:6–14. [HT]
- Wlodarski, Rafael, John Manning, and Robin I. M. Dunbar. 2015. Stay or stray? evidence for alternative mating strategy phenotypes in both men and women. *Biology Letters* 11(2):20140977.
- Wrangham, Richard W. 1982. Mutualism, kinship and social evolution. In *Current problems in sociobiology*. King's College Sociobiology Group, eds. Pp. 269–289. Cambridge: Cambridge University Press.
- Wright, Anthony A., Jacquelyne J. Rivera, Stewart H. Hulse, Melissa Shyan, and Julie J. Neiwirth. 2000. Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General* 129(3):291–307.
- Wu, H. X., Y. Li, B. H.-H. Ching, and T. T. Chen. 2023. You are how you speak: the roles of vocal pitch and semantic cues in shaping social perceptions. *Perception* 52(1):40–55, <https://doi.org/10.1177/03010066221135472>. [PES/YO]
- Zahavi, Amotz, and Avishag Zahavi. 1999. *The handicap principle: a missing piece of Darwin's puzzle*. Oxford: Oxford University Press.
- Zarate, J. M. 2013. The neural control of singing. *Frontiers in Human Neuroscience* 7:237. [ARH]