

COMPARATIVE COGNITION & BEHAVIOR REVIEWS

The Links Between Pitch, Timbre, Musicality, and Social Bonding From Cross-Species Research

Bernhard Wagner^{1,2}, Marisa Hoeschele^{1,2,3}

¹*Acoustics Research Institute*, ²*Austrian Academy of Sciences*, ³*University of Vienna*

Music has been proposed as a “coevolved system for social bonding” (Savage et al., 2020), emphasizing a role of biology in human music. Cross-species studies can help us gain insight into the phylogenetic, behavioral, and/or physiological constraints of the biological abilities underlying music, known as musicality. Cross-species research of past years has focused largely on rhythmic abilities and vocal learning in nonhuman species. By surveying the existing literature, this review shows that spectral aspects of musicality—such as octave equivalence, consonance preference, and saliency of timbre—remain understudied, yet appear crucial to social bonding. We delineate how these abilities may facilitate social bonding and propose the four biological traits of vocal learning, vocalizations with clear harmonics, differing vocal ranges, and simultaneous vocalizing, which primarily appear to constrain the abilities. We illustrate how these traits interact in shaping musicality and make suggestions and predictions regarding future research on the connection between spectral musicality traits and social bonding.

Keywords: musicality, vocal learning, octave equivalence, consonance, timbre

Introduction

It has long been suggested that there may be biological roots to human music. Archaeological evidence, research on infants, and cross-species research on nonhuman animals all support this idea (e.g., Bowling et al., 2017; Brown & Jordania, 2013; Hoeschele et al., 2015; Savage et al., 2020). However, many competing hypotheses propose how and why human musical behavior may have an evolutionary component.

A main problem in the study of musical behavior evolution in humans is posed by confounds of culture and biology. Here, cross-species studies present a way to avoid these confounds; in nonhuman animals, the degree of enculturation to human music can be experimentally controlled. As such, cross-species studies of musicality

can lead us to important insights on potential biological constraints on human musicality. We define *musicality* as the sum of the abilities that enable but do not constitute the cultural construct that is human music (Hoeschele, 2017). These abilities may also be found in nonhuman animals. For example, discrimination and generalization tasks can determine whether nonhuman species perceive and categorize sounds in similar ways to humans, whereas preference paradigms can give insight into the sounds to which they show an attraction. A consideration of the ecological relevance of the studied traits and abilities for each species allows us to come to a better understanding of how environment and behavior may shape them. Further, we can compare whether the identified constraints consistently appear to shape musicality in different species—that is, whether musical abilities are cases of analogous or

homologous evolution. As such, cross-species research allows for conclusions on the evolutionary origins as well as phylogeny of musical abilities.

Much attention has been paid lately to the rhythmic abilities of several species (e.g., Celma-Miralles & Toro, 2018; Cook et al., 2013; Hattori & Tomonaga, 2019; Honing et al., 2012, 2018; Patel et al., 2009). An important area of study involves the vocal-learning beat synchronization hypothesis (Patel, 2006; Patel et al., 2009; Schachner et al., 2009). The original version of this hypothesis suggested that the ability to synchronize with a musical beat was inextricably linked to vocal learning. Although this concept has since been refuted (Cook et al., 2013), the research connected to the idea that vocal learning and rhythmic synchronization are related remains valuable and highly important. However, our review shows that the existing literature suggests biological components of other (nonrhythmic) aspects of musicality, namely, pitch and timbre. These aspects—more specifically, perception of octave equivalence, attraction to consonance, and salience of timbre—remain understudied in nonhuman animals. We hypothesize that four traits in our vocal behavior shape the musical properties of spectral features in human musicality: (a) vocal learning, (b) harmonic clarity in vocalizations, (c) differing vocal ranges, and (d) simultaneous vocalization and duetting. We explain these traits and how they connect to three important aspects of musicality (viz., octave equivalence, consonance, and salience of timbre). The fact that the four traits appear in different species to different degrees allows us to suggest possibilities for new research and to test our hypotheses of how they may shape musicality. Subsequently, we consider these aspects in terms of their connection to the “music and social bonding hypothesis” (Savage et al., 2020, Abstract): the idea that musical abilities in humans evolved to facilitate social bonding.

Author Note: Bernhard Wagner, Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040, Vienna, Austria.

Correspondence concerning this article should be addressed to Bernhard Wagner at bwagner@kfs.oeaw.at

Acknowledgments: XXX

The Four Traits

Vocal Learning

Humans are one of the minority species that require experience to produce normal vocalizations. How this ability relates to human vocal behavior is clear: We need this ability not only to learn new songs but also to acquire language. Species whose vocal learning abilities are unequivocal are humans, elephants, bats, several pinniped species, whales, dolphins, oscine songbirds, hummingbirds, and parrots (Reichmuth & Casey 2014; Tyack, 2008). However, it is important to mention recent propositions that vocal learning should not be seen as a binary trait but as a modular concept (see, e.g., Jarvis, 2019; Wirthlin et al., 2019), with intermediary abilities such as subtle changes to innate vocalizations based on experience.

This is especially important, as parallels between vocal learning in humans and other species have been a focus of much musicality related research over the years (see Hoeschele et al., 2015, for a review), perhaps overlooking musicality traits in species with more subtle vocal modification abilities. In addition, the overarching focus on vocal learning has seemingly passed over the importance of or not explicitly considering the relevance of the other three traits. As we discuss next, vocal learning may enhance the effects of the other traits, which may be why these three traits were initially overlooked.

Harmonic Clarity in Vocalizations

Vocal fold vibration in humans, as in many other species, produces harmonic sounds. This means that human vocalizations contain a fundamental frequency (the percept of which is pitch) and harmonics that occur at integer multiples of this fundamental frequency. Human adults, unlike other species and human babies, avoid nonlinearities (which would occlude this harmonic structure) in speech as in song (Arnal et al., 2015; Fitch et al., 2002). As such, humans have particularly clear harmonics in their vocalizations, making them a salient aspect of musical sounds. Many animals have a higher amount of noisy vocalization in their repertoire (Fitch et al., 2002). As we discuss, this lack of harmonic clarity may be detrimental to the development of consonance preference and octave equivalence. The reason for this is that the salience of pitch and human preference to consonance may have their roots in the harmonic clarity within the human voice (Bowling & Purves, 2015; Bowling et al., 2018).

Differing Vocal Ranges

Human adults have differing vocal ranges, in part related to their sex. Adult vocal ranges are also usually lower than children's. This is especially true for adult male voices. Because human vocal ranges span multiple octaves—something that is more common in larger animals—octave equivalence can help us vocalize together. More generally, the same applies for salience of consonance, as we explain. Sexual dimorphism of vocal range is unusually pronounced in humans (Puts et al., 2016). As such, the more common differences between adults' and juveniles' vocal pitch because of size difference during development are perhaps the most promising for comparative research.

Simultaneous Vocalization and Duetting

Humans sing simultaneously or in duets (turn-taking) and communicate with carefully timed vocal exchanges. Matching harmonic information, in both vocal learners and vocal nonlearners, may be important during many of these vocal exchanges. Importantly, Wirthlin et al. (2019) defined such behaviors (under the term *vocal coordination*) as a constituent ability of their modular concept of vocal learning. We recognize this important perspective and stress that vocal behavior abilities may be interconnected in more ways than we can discuss here while remaining focused on human musicality.

The following sections show how these four traits interact and constrain three musicality-related abilities in humans, namely, octave equivalence, consonance preference, and salience of timbre.

Octave Equivalence

Octave equivalence describes the perception that notes separated by an octave (i.e., a doubling in frequency; Burns, 1999; Patel, 2003), sound alike—in some instances even more alike than that note and a note with a frequency closer to its own (see Allen, 1967; Hoeschele et al., 2012; Kallman, 1982). Although cultural differences in pitch perception and musical scales exist (Hove et al., 2009), all musical cultures use the octave (Burns, 1999; but see Jacoby et al., 2019, discussed in more detail next). Evidence also suggests that pitch is mapped adhering to octave circularity in the auditory thalamus (Langner & Ochse, 2006). Thus, it is unsurprising that even infants at a preverbal stage were shown to be less responsive to tones with a fundamental frequency an octave above a sound with which they were familiar in comparison with other tones, suggesting that habituation to a tone was generalized

across octaves (Demany & Armand, 1984).

Once vocal learning begins, human children use octave equivalence for successful imitation. For example, when a young child imitates their father singing a melody with a fundamental frequency outside the child's vocal range, the child spontaneously produces the melody transposed by an octave, which we perceive as an accurate imitation (see Hoeschele, 2017, for a review). Children use octave equivalence when they are presented with words spoken at a fundamental frequency below their vocal range even if this use of octave equivalence results in an imitation still lower than their usual speaking pitch (Peter et al., 2008). Thus, it appears that octaves sound similar to one another for children (Peter et al., 2008, 2009) as well as for adults (Peter et al., 2015). The harmonic physical structure of the human voice allows for an explanation of this phenomenon: If one cannot reproduce the actual fundamental frequency of a sound, then producing its octave results in an overlap of half the original note's harmonics and thereby the physically most similar sound that is achievable (Hoeschele, 2017).

From the preceding research about human vocal learning, we might also conclude that vocal learning is highly relevant in octave equivalence. This idea can be tested by looking at whether, in other species, vocal learning and perception of octave equivalence appear to be related. However, the little research that has been conducted in this vein is contradictory (Burns, 1999; Hoeschele, 2017; Hoeschele et al., 2015) and often nonstandardized. So far, two studies have suggested octave equivalence perception in nonvocal learning species: rhesus monkeys (*Macacca mulatta*; Wright et al., 2000) and rats (*Rattus norvegicus*; Blackwell & Schlossberg, 1943). However, we note that the former study has a sample size of only two individuals and the latter has been criticized as not controlling for harmonics, meaning that results can be explained without invoking octave equivalence (see Burns, 1999). Among vocal learning species, one bottle-nosed dolphin (*Tursiops truncatus*) spontaneously octave-transposed sounds outside her preferred vocal range (Richards et al., 1984), and European Starlings failed to show octave equivalence (Cynx, 1993) in a procedure later shown to also not show octave equivalence in humans (Hoeschele et al., 2012). Black-capped chickadees (*Poecile atricapillus*; Hoeschele et al., 2013) as well as budgerigars (Wagner et al., 2019) failed verified and standardized tests of octave equivalence first conducted in humans (Hoeschele et al., 2012). As such, it becomes clear that vocal learning in itself does not appear to require octave equivalence. The next thing we considered is that octave equivalence is

useful for vocal learning only if vocal learning happens in contexts where the learner is unable to imitate a template sound as originally presented, because the template sound's fundamental frequency lies outside the learner's vocal range. Logically this would occur in species where vocal ranges differ—between adults and children as well as between sexes.

Martins et al. (2009) described how the male voice drops roughly an octave from childhood to adulthood, reiterating the octave relationship between male adult and child voices. As such, and in addition to the findings from Peter et al. (2008, 2009, 2015), it appears that vocal range difference between adults and children may be an important constraining factor on octave equivalence, especially considering that childhood is when most of human vocal learning occurs. This finding is especially relevant to cross-species research, as differences in fundamental frequency between males and females are unusually pronounced in humans compared with other apes, especially considering that body size in humans is less sexually dimorphic than in other simian species (Puts et al., 2016). As such, the relationship between male and female vocal pitch may not translate readily from humans to other species. Meanwhile, pronounced pitch differences between adult and juvenile animals are common in many animals (see Matrosova et al., 2007). As such, even vocal nonlearning animals may benefit from perceiving octave equivalence to facilitate communication between fully grown adults and juveniles whose vocalizations are still higher in pitch because of their smaller size. In such a case, some mechanism of generalizing a vocalization having the same “meaning” or “message” when produced at a different pitch would be needed. Octave equivalence could thus be a suitable mechanism to facilitate recognition of vocalization types at different pitches for both vocal learners and vocal nonlearners. Potentially this could also be facilitated by perception of relative pitch, that is, the relative relationships between fundamental frequencies in a series of vocalizations instead of paying attention to absolute frequencies. However, relative pitch appears to come more easily to humans than to nonhuman animals, perhaps because it is rather abstract and requires recognition of patterns over time (see Hoeschele, 2017, for a review). As human children use harmonic structure in the form of octave matching when learning to speak and as nonhuman primates have also been shown to do some pitch matching (Sugiura, 1998), there may be similarities in attendance to the harmonic series among primates. However, these could be shared with more distantly related species as well. If so, this attendance to the harmonic series may be constrained

to mammals, as mammals have longer developmental phases than, say, birds. This may increase mammals' need for a mechanism to generalize subadults' higher pitched calls. As such, octave equivalence may be more common in mammals, which is supported by the literature so far (Hoeschele et al., 2013; Richards et al., 1984; Wagner et al., 2019; Wright et al., 2000) as reviewed earlier.

Octave equivalence being more common in mammals appears even more likely when we consider harmonic vocal structure. Many bird vocalizations have no harmonic information beyond the fundamental frequency (Harma & Somervuo, 2004). Such vocalizations would not constrain octave equivalence perception. Where they do produce harmonic sounds, birds' high-frequency vocal ranges have harmonics that are spread much further apart in absolute frequency (being interval multiples of the fundamental). As such, vocalizations require a larger frequency range to span multiple octaves, which may also play a role in birds perceiving that they are less salient. This could be one reason why no bird species has been found to perceive octave equivalence, with two species testing negative in a standardized test for octave equivalence (Hoeschele et al., 2013; Wagner et al., 2019). Some animals may generally pay more attention to overall spectral shape and less to individual frequency bands within a sound's spectrum, as did European starlings in a study by Bregman et al. (2016). In this context, it is interesting that the results from both Hoeschele et al. (2013) and Wagner et al. (2019) suggest that the tested bird species may group tones in a nonrandom way different from octave equivalence. Both species responded significantly less to octave five (262–494 Hz) notes that they had been trained to respond to in octave four (523–988 Hz). This reaction is the opposite of what would be expected from a species that perceives octave equivalence. Thus, we should acknowledge that there appear to be other ways of enabling successful imitation of sounds outside an individual's vocal range besides octave equivalence. For example, budgerigars, when imitating human voices that lie outside their vocal range, reproduce a formant frequency in the human voice that is close in pitch to frequencies at which they usually vocalize (Sclan, 1999). As such, it is important for researchers to remember that the constraints important to octave equivalence (e.g., imitation outside one's vocal range) could also result in the development of different mechanisms that fulfill the same functions (e.g., formant imitation).

Although it is common that juvenile individuals have higher pitched vocalizations compared with adults because of their body size (Matrosova et al., 2007), the picture is more muddled for sex differences. Puts et al.

(2016) have shown how vocal sexual dimorphism in pitch is unusually pronounced in our species compared with phylogenetically close groups. Research on humans has often stated that male and female vocal pitch are an octave apart, on average (see, e.g., Savage et al., 2020), which strongly suggests a connection to octave equivalence either to facilitate vocal learning between males and females or to facilitate harmonic overlap when singing together. However, that information could be outdated. Titze (1989) is often cited as a source, yet although that article provides an interesting mechanistic model of human vocal anatomy, the data concerning average male and female pitch used therein are taken from Kent (1976), who compiled data from numerous studies from 1906 to 1975. Research has shown that the speaking pitch of young women has significantly lowered from the years 1945 to 1993 (Pemberton et al., 1998). A recent large sample study by Berg et al. (2017) found that female speaking pitch was “six to seven semitones lower than previously described” (p. XX), with the male’s conversational voice at 111.9 Hz (corresponding most closely to the Western music note A2) and the female’s at 168.5 Hz (corresponding most closely to the Western music note E3) instead of the previously reported ~220 Hz (Kent, 1976).

This result is consistent with findings that show a confound of cultural gender norms and speaking pitch. For example, Japanese women show higher fundamental frequency than Dutch women (van Bezooijen, 1995) corresponding with cultural gender norms. Furthermore, males who perceive themselves as more feminine were shown to have higher average fundamental frequency than men who perceive themselves as more masculine (Weirich & Simpson, 2018). However, it is important to note that these studies (like the oft-cited Titze, 1989) describe speaking pitch. General vocal range while singing may still show a different picture. For example, a study of approximately 80 members of a nonprofessional choir found that the average lowest and highest pitch in males and females was about an octave apart (Hacki, 1999). Still, choir members would be biased to train to fit in the musical categories of a choir that may favor octave differences because of musical harmony, an apparent confound.

As such, the final of our four proposed traits, simultaneous vocalization (and duetting), should be considered. It is easy to see how having differing vocal ranges within a species interplays with singing together. When children sing with their parents or men and women sing together, individuals will at times not be able produce the same pitch. However, when they instead produce the octave of a given tone, as many harmonics as possible between the

two distinct tones will overlap. In this light, we wish to briefly discuss an important study by Jacoby et al. (2019). The authors found that native Amazonian Tsimané naïve to Western music did not use the octave to copy notes outside their range in a singing task in which the Western subjects did. This finding has been interpreted as a contradiction to the idea that octave equivalence is universal across cultures. However, in a later study by the same lab (McPherson et al., 2020), Tsimané were more likely to perceive two simultaneously played notes as one if they were separated by an octave. This finding supports the importance of the octave being used in humans singing together creating a perceptual fusion of the composite voices. Such perceptual fusion can also be achieved with notes separated by other small integer frequency ratios. These ratios are known in Western music theory as *consonant intervals*.

Consonance

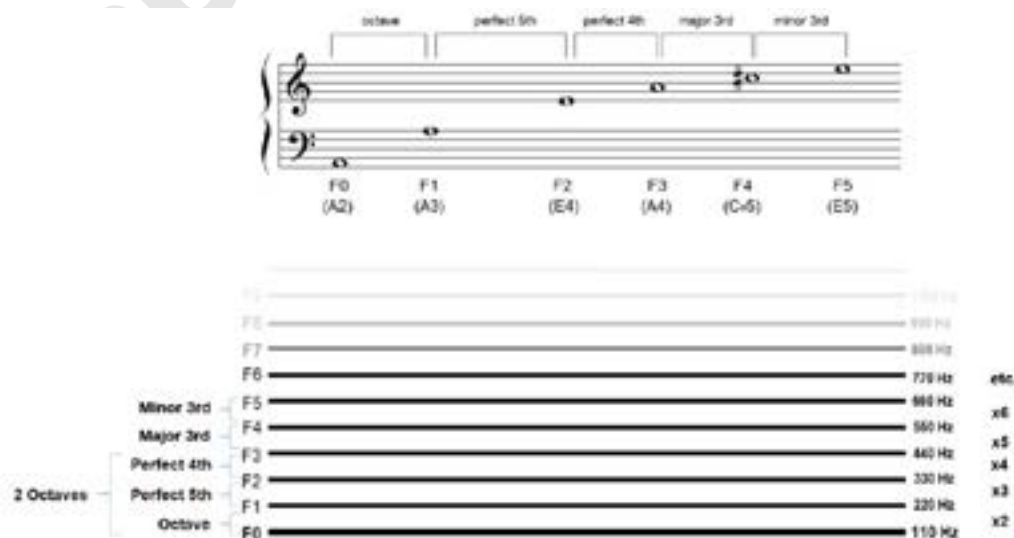
Consonance is a term from Western music theory describing combinations of tones that are considered to be perceptually pleasant (see, e.g., Bowling & Purves, 2015; Krumhansl, 1990; Terhardt, 1984), whereas tone combinations that are considered unpleasant are referred to as *dissonant*. For centuries, individuals have observed that the defining characteristic of these pleasant tone combinations seems to be small integer ratios between the constituting frequencies (Chen, 1996; Crocker, 1963). This phenomenon is not restricted to Western music, as intervals with small ratios are considered consonant by listeners from a variety of cultures and are found throughout music from around the world (Burns, 1999). Several infant studies have suggested that a preference for consonance appears to develop early in life (Masataka, 2006; Perani et al., 2010; Schellenberg & Trehub, 1996; Trainor & Heinmiller 1998; Trainor et al., 2002; Trehub, 2003; Zentner & Kagan 1996, 1998). Yet some cross-cultural studies contradict the universality of consonance (see, e.g., Athanasopoulos et al., 2021; Jacoby et al., 2019). Additionally, research has shown that familiarity with consonant sounds (Platinga & Trehub, 2014) and with music theory (see, e.g., Bowling et al., 2018; McDermott et al., 2010) seem to play an important role in the assessment of consonance. As such, an ongoing debate still exists about why exactly tone combinations with small integer ratios are perceived as more pleasant than others (see, e.g., Bowling et al., 2017; McDermott et al., 2016; Virtala & Tervaniemi, 2016). Recent work has emphasized the potential role of the biological and physical properties of the human voice

in consonance assessment (e.g., Bowling & Purves, 2015; Schwartz et al., 2003; Terhardt, 1984). Human preference for consonant over dissonant intervals correlates with a preference for harmonic sounds such as those found in the human voice (Bowling et al., 2018; Cousineau et al., 2012; McDermott et al., 2010). Because harmonic sounds comprise frequencies that are integer multiples of one another, they prominently feature the most consonant intervals (beginning with the octave 2:1). As such, it could be that a preference for natural human vocalizations leads to consonance preference in humans. In fact, the most prominent intervals in vocalizations (harmonics closest to the fundamental) are also the intervals with the highest cross-cultural ratings for consonance. See Figure 1 for a visualization.

The hypothesis that attraction to harmonic vocalizations is the biological foundation of preference to consonance has been dubbed the “vocal similarity hypothesis” (Bowling & Purves, 2015; Bowling et al., 2018). This hypothesis invites cross-species comparison: If the basis of human consonance preference lies in the physical properties of the human voice, then we should expect to also find such a preference in other animals with vocal output that is similarly harmonic to that of humans. However, research with animals in this direction remains limited, with few species having been studied and studies lacking comparability because of the use of differing

paradigms. Some studies have tested only for the ability to discriminate between consonant and dissonant stimuli and not for consonance preferences. Such studies found that Japanese macaques (*Macacca fuscata*; Izumi, 2000), European starlings (*Sturnus vulgaris*; Hulse et al., 1995), and Java sparrows (*Lonchura oryzivora*; Watanabe et al., 2005) are able to discriminate between consonant and dissonant stimuli. Pigeons’ ability to discriminate complex chords suggests they would also be able to discriminate consonance from dissonance (Brooks & Cook, 2010; Porter & Neuringer, 1984; see Toro & Crespo, 2017, for a review). However, although the ability to discriminate these stimuli is a evident prerequisite for consonance preference, no definitive sign for preference can be inferred here. Results were mixed in the studies that have directly tested consonance preference: In Tungara frogs (*Physalaema pustulosus*), manipulation of frequency ratio in mating calls (producing consonant or dissonant intervals) did not influence attraction (Akre et al., 2014). In an early study, Fannin and Braud (1971) found that albino rats (*Rattus norvegicus*) preferred consonance. However, a more recent study found that brown rats (*Rattus norvegicus*) learned to discriminate between consonant and dissonant chord only in the stimuli with which they were trained but failed to generalize the discrimination pattern to novel stimuli (Crespo-Bojorque & Toro, 2015). This result suggests that the rats were not able to categorically

Figure 1. The frequency bands (F0–F6 demarcating fundamental frequency to the sixth frequency band above it) of harmonic sounds prominently feature consonant intervals. The fundamental (i.e., the first harmonic) and the second harmonic compose a perfect octave (2:1), the second and the third harmonic a perfect fifth (3:2), and the third and the fourth harmonic a perfect fourth (4:3): All of these intervals are considered to be consonant cross-culturally (Burns, 1999). In addition, the fourth and the fifth harmonic compose a major third (5:4), and the fifth and the sixth harmonic a minor third (6:5), which are also considered consonant in Western music theory (albeit to a lesser degree than the aforementioned “perfect” intervals). These relations are also presented in musical notation starting with A2, the corresponding note name for 110 Hz in Western music notation. Notes diverge minimally from just temperament tuning; the note names given are the closest corresponding just temperament note names.



distinguish consonance and dissonance and relied on memory of specific chords to solve the task. Cotton-top tamarins (*Sanguinus oedipus*) showed no preference for consonant over dissonant stimuli in a study by McDermott and Hauser (2004). Campbell's monkeys (*Cercopithecus campbelli*) also showed no preference in a study by Koda et al. (2013). One study found that an infant chimpanzee (*Pan troglodytes*) preferred consonant intervals (Sugimoto et al., 2010). However, Chiandetti and Vallortigara (2011) pointed out that the chimpanzee's early auditory environment had not been controlled. As the individual was hand-raised by humans, it may well have been exposed to music before it was tested. As such, the preference for consonance may have arisen from familiarity rather than a natural preference for consonant stimuli. Even less research has been conducted with birds, yet the few results suggest that some species may in fact be attracted to consonance. Songs of the musician wren (*Cyphorhinus arada*) and the hermit thrush (*Catharus guttatus*) contain sequences of harmonic intervals (Doolittle & Brumm, 2012; Doolittle et al., 2014). Production of harmonic intervals sequences in the great tit (*Parus major*) is related to mating success as shown by Richner (2016), which suggests that a preference for consonance may be driving this success. In a study by Chiandetti and Vallortigara (2011), newly hatched chicks (*Gallus domesticus*) that had been incubated in acoustic isolation were imprinted on a red plastic cylinder—meaning they would perceive the object to be their mother. In a behavioral experiment, these chicks were then allowed to freely approach either of two red cylinders identical to the one on which they had imprinted. One of these cylinders was associated with playback of consonant music and the other of dissonant music. The results showed that the chicks preferentially spent time with the cylinder associated with consonance rather than the one associated with dissonance. The study's authors suggested that the chicks may have used consonance as a cue to which object was more likely to be their mother. Just like human harmonic vocalizations, a mother hen's harmonic vocalizations prominently feature consonant intervals, whereas sounds from inanimate objects are mostly nonharmonic (Chiandetti & Vallortigara, 2011). As such this study provides evidence of a species being attracted to consonance in an ecologically relevant setting. Attraction to harmonic sounds may facilitate the distinction between conspecific's sounds (which are usually harmonic) and environmental sounds (which often are not harmonic; Chiandetti & Vallortigara, 2011). Thus, it can be seen how clear harmonic vocalizations constrain consonance preference.

With regard to vocal learning, our group hypothesized that a vocal learning species may especially benefit from a perceptual frame favoring harmonic vocalizations (and thereby consonance). Because potential tutors would produce mainly harmonic sounds such a preference could facilitate imitating more relevant sounds from tutors and less irrelevant environmental sounds. However, adult budgerigars (*Melopsittacus undulatus*), a vocal learning species, did not show preference for consonance in a place preference paradigm (originally developed by Hoeschele & Bowling, 2016), which let them choose between being close to a sound source playing consonant or another sound source playing dissonant sound stimuli (Wagner et al., 2020). As such, evidence for heightened attraction to consonance cannot be supported by this study. However, budgerigars not being attracted to consonance may not be out of line with the vocal similarity hypothesis. The rationale for this connects back to the importance of harmonic vocalizations: Although budgerigar vocalizations do contain harmonics, their vocal communication features a relatively large proportion of nonlinear phenomena (Lavenex, 1999; Tu et al., 2011). Nonlinear phenomena in vocalizations generate deterministic chaos, which masks harmonic structure with energy similar to turbulent noise (Fitch et al., 2002). Therefore, the harmonic structure within such vocalizations tends to be obscured. It bears repeating that vocalizations with nonlinearities are usually avoided in speech as in song by human adults (Arnal et al., 2015; Fitch et al., 2002). As nonlinearities are common even in the budgerigars' learned contact call (Lavenex, 1999; Tu et al., 2011), this species may not benefit from attending to harmonic structure when imitating vocalizations, and thereby may not benefit from favoring consonance. This emphasizes the importance of vocalizations with clear harmonics. However, only more research can show whether and how vocal learning may constrain consonance preference.

Recent research suggests that although the “pleasantness” effect of consonance may not be as cross-culturally shared as previously assumed (Athanasopoulos et al., 2021; Jacoby et al., 2019), consonant intervals cross-culturally have the quality of sounding less like two separate sounds than dissonant intervals do (McPherson, Dolan, & Durango, 2020). This is logical considering that consonant intervals also exist in the overtones of harmonic sounds, creating large overlap in harmonic structure. It is also highly relevant to the remaining two traits of the four: different vocal ranges and vocalizing together. As with the octave, simultaneous vocalization using consonant intervals can make two voices meld in our perception facilitating social

bonding. Humans having different vocal ranges makes it more important to pay attention to harmonic information and perhaps encourages vocalizing at consonant intervals to create perceptual fusion. It is worth mentioning at this point that the results from Berg et al. (2017) that showed that male and female speaking voices appear to no longer be an octave apart do not as such contradict the idea that human vocal anatomy could be an “anatomical adaptation for vocal harmonization” as put forth by Savage et al. (2020). However, the results from Berg et al. may suggest that the harmonization would not be an octave as Savage suggested but a perfect fifth, the second most consonant interval. This is in line with the finding by Peter et al. (2015) that women imitate men at an interval of a perfect fifth. As such, human singing groups may be predisposed to harmonize at different consonant intervals. This can create the impression of sounds being merged into being perceived as one louder, larger sound with richer timbre. Thus, what we normally refer to as harmony—the production of more than one pitch simultaneously—can alter the perceived timbre or quality of a sound.

Timbre

The concept of timbre is perhaps most easily explained by considering it to be what constitutes the difference between the same note played by, for example, a piano and a violin and the same note sung by two people. For humans, the saliency of pitch makes timbre easily separable from pitch in many cases. This is facilitated by the harmonic structure of the human voice, where (slightly simplified) we perceive the fundamental frequency as pitch and the other spectrotemporal features of the sound as timbre. Because harmonic sounds are typically automatically evaluated in this way by our species, we can learn to sing a melody that is not sung to us but played on a piano. However, this ability has limitations: Humans are able to detect only the pitch of a simple tune or the melody of a piece of music, but even experts have trouble identifying the pitches of all the accompanying instruments and harmonizing vocals, for example (e.g., Klapuri & Davy, 2006). Further, nonmusicians have trouble identifying whether two notes that differ in timbre also differ in pitch (Pitt, 1994). This suggests that pitch and timbre are not always clearly separable when we have a complex harmonic sound, which is especially true if the timbre cannot be parsed easily. Familiarity with a particular timbre can also affect the difficulty of pitch identification. For example, musicians with absolute pitch more quickly identified the key of a piece when it was performed with

their primary instrument (Marvin & Brinkman, 2000). These results suggest that successful pitch identification depends on both the timbral and harmonic context. This is where clearly harmonic vocalizations, differing vocal ranges, and simultaneous vocalization are relevant as, perceptually, the lines between harmony and timbre may blur: Think of how in a Gregorian chant many voices singing a melody in unison merge to the effect that individual voices become difficult to distinguish. In a way, this merged group of voices have a timbre that is different from that of each individual one voice alone. Now consider that when some of the voices add harmony, there is more than one fundamental frequency at the same time. Some of these fundamental frequencies may be perceived as part of the timbre, and similarly, some of the timbre of one note in the harmony may be perceived as part of the pitch of another note. Especially if many or even all of the frequency bands overlap, harmony and timbre may blend.

To date, there has been little cross-species study of timbre perception in animals, but there is evidence that birds do not separate pitch and timbre the way humans do (Bregman et al., 2016, Hoeschele et al., 2014). More research is needed, and in the following sections we attempt to provide promising directions for that study.

What Conclusions Can We Draw About the Mechanisms Underlying Musicality From Cross-Species Musicality Research?

In sum, research supports our hypothesis that four human vocal abilities are important in the spectral aspects of musicality: vocal learning, harmonic clarity in vocalizations, differing vocal ranges, and simultaneous vocalization and duetting.

Table 1 displays an overview of the species that have been studied regarding (a) their ability to discriminate consonant and dissonant stimuli as well as (b) their ability to transfer this to novel stimuli, (c) their preference for consonance versus dissonance, (d) the production of consonant intervals in their vocalizations, (e) the perception of octave equivalence, and (f) the perception of salience of timbre. Whether these species are vocal learning species is known and represented in the table. However, the harmonic clarity of the vocalizations and whether they span multiple octaves or whether different individuals have different vocal ranges have not been studied methodically with regards to musicality. If our hypothesis about the four traits is correct, these additional features could be highly relevant to determining why other species share our perceptual and vocal abilities.

Table 1. Animals Tested in Pitch-Related Musicality Experiments and Traits They Possess.

Species	Consonance Distinction	Consonance Transfer	Consonance Preference	Consonance Production	Octave Equivalence	Timbre Salience	Vocal Learning
Mammals							
Humans	√	√	√	√	√	√	√
Rats	√	x	?	x	?	—	x
Cotton-top tamarins	—	—	x	x	—	—	x
Campbell's monkeys	—	—	x	x	—	—	x
Chimpanzee	—	—	?	x	—	—	x
Rhesus macaques	—	—	—	x	√	—	x
Japanese macaques	√	√	—	x	—	—	x
Bottle-nosed dolphin	—	—	—	x	√	—	√
Birds							
Musician wren	—	—	?	√	—	—	√
Hermit thrush	—	—	?	√	—	—	√
Great tit	—	—	√	√	—	—	√
Black-capped chickadee	√	√	—	—	—	√	√
European starling	√	√	—	—	?	—	√
Budgerigar	—	—	x	—	x	—	√
Java sparrow	√	√	—	—	—	—	√
Pigeon	√	x	—	—	—	—	X
Domestic chicks	√	—	√	—	—	—	x

Note: A check mark signifies that a species has tested positive for the ability; an X mark signifies that a species tested negative; a question mark signifies that results are not entirely conclusive; a dash signifies that there are no data to date. The references for the respective species are as follows: rat (Blackwell & Schlosberg, 1943; Crespo-Bojorque & Toro, 2015; Fannin & Braud, 1971), cotton-top tamarin (McDermott & Hauser, 2004), Campbell's monkey (Koda et al., 2013), chimpanzee (Sugimoto et al., 2010), rhesus macaque (Wright et al., 2000), Japanese macaque (Izumi, 2000), bottle-nosed dolphin (Richards et al., 1984), musician wren (Doolittle & Brumm, 2012), hermit thrush (Doolittle et al., 2014), Great tit (Richner, 2016), black-capped chickadee (Hoeschele et al., 2013, 2014), European starlings (Bregman et al., 2016; Cynx, 1993; Hulse et al., 1995), budgerigars (Wagner et al., 2019), Java sparrows (Watanabe et al., 2005), pigeons (Brooks & Cook, 2010), and chickens (Chiandetti & Vallortigara, 2011).

Directions for Future Research and Their Relationship to the Social Bonding Hypothesis

Table 1 shows clearly that so far no one species has been tested for all of these traits and abilities. Having at least one such species would give us better insight into the validity of our hypothesis. Ideally, focusing on several species with and without vocal learning abilities, differing harmonic clarity, with and without differing vocal ranges, with a variety of vocal performance styles (e.g., simultaneous vocalization and/or duetting), and with and without vocal learning abilities would be ideal to allow for a more comprehensive understanding of the relevance of these traits for musicality. As such, further research is needed to test the ideas and hypotheses presented in this review. One objective of such research should be to expand the number of species tested for octave equivalence, consonance preference, and integration of timbre. Testing phylogenetically distant groups would provide a valuable

big-picture perspective with regards to the evolutionary paths of given traits. This view is especially important, as the four traits we emphasize may develop analogously in different clades and thus cannot be assumed to be constrained by phylogeny. Phylogeny may also be important in cases where species with a common ancestor that, for example, duetted or had differing vocal ranges may show aspects of spectral musicality even if one of the descendant species did no longer show this behavior. As mostly birds and mammals have been studied with regards to spectral musicality, anurans and insects can be used as outgroups for comparison (e.g., Hoeschele et al., 2017).

Absence of evidence does not equal evidence of absence, so research in this field can be difficult because of the risk of negative results. It is therefore a welcome development that registered reports are appearing more often in many journals, as negative results in this field are often of great value. In addition to registered reports, the risk of negative results can also be minimized by taking

care to design studies that are ecologically relevant to the tested species in paradigm (such as Chiandetti & Vallortigara, 2011) and in stimuli (Snowdon & Teie, 2010; but see Wagner et al., 2020). To this end, one should consider how the subject species may benefit from the studied musicality traits and which circumstances may constrain development of such traits in certain species. Additionally, the risk can be minimized by using comparative procedures across multiple species, especially humans, such that there are always positive results to use as a comparison and the different species are directly comparable. Several suggestions and predictions regarding future research in the areas of octave equivalence, consonance preference, and saliency of timbre can be inferred from the literature reviewed here, which we describe in detail next.

Octave Equivalence

Given the studies reviewed here, testing more species for octave equivalence while considering the four traits appears to be a promising endeavor. This could be achieved, for example, by assessing the pitch range changes from juvenile to adult stages of development and prioritizing vocal learning and vocal matching species. We would expect octave equivalence to be most prevalent in larger vocal learning animals with slow juvenile development and low vocalizations (e.g., elephants or baleen whales) and expect it to be less common in smaller animals with short juvenile development and high-pitched vocalizations, such as many bird species.

It would also be interesting to study adult–juvenile vocal communication directly. In this review we considered the possibility that different body sizes and consequently different pitches in juvenile–adult communication may facilitate development of octave equivalence to allow generalization of calls at different pitches. This study is even more interesting considering findings that vocal nonlearners can adjust some frequencies in their calls, especially when call-matching vocalizations produced by conspecifics. For example, this occurs with “coo” calls of Japanese macaques (Sugiura, 1998), who are close relatives of rhesus macaques. The latter species has been suggested to perceive octave equivalence (Wright et al., 2000). If rhesus macaques do similar vocal matching of “coo” calls, a compelling study would be whether juveniles, whose calls are still at a higher pitch because of their smaller size and developmental stage (Hammerschmidt et al., 2000), use octave equivalence in call matching. For example, they may match the octave of frequencies that they cannot produce because of their smaller size. By the same line of reasoning, testing Japanese macaques for

octave equivalence could also prove a fruitful endeavor. Generally speaking, call matching could be relevant to musicality and may be a rewarding place to look for octave equivalence production in other species. This also relates back to the idea that call matching can be seen as an intermediary form of vocal learning (between vocal learning and nonlearning; Wirthlin et al., 2019). As such, considering musicality studies in species that are not known to engage in simultaneous vocalization or “hard” vocal learning but engage in call matching will significantly enlarge comparative possibilities.

When testing for the perception of octave equivalence, the most promising method appears to be the verified standardized procedure by Hoeschele et al. (2012, 2013; see also Wagner et al., 2019), as it can be implemented with a variety of species retaining comparability. However, it would be at least as interesting to derive a paradigm of octave equivalence production from the spontaneous occurrence of octave transposal in Richards et al. (1984). Such a paradigm could train vocal mimicking animals to imitate sounds (as was done successfully with orcas [*Orcinus orca*] in Zamorano-Abramson et al., 2018) and then test whether they spontaneously octave transpose target sounds that are outside their vocal range (Wagner & Hoeschele, 2022). This test would also minimize the risk of negative results with animals potentially employing alternative strategies of mimicry as (e.g., as we discussed earlier) budgerigars do (Sclan, 1999).

Consonance

This review describes how vocal learning may be relevant to consonance preference because a vocal learning species may benefit from heightened attention to harmonic sounds and thereby to musical consonance. Future consonance-related cross-species research may also want to consider the general importance of harmonic vocalizations to the emergence of consonance preference because harmonicity in a species’ vocalizations may correlate directly with their attraction to consonance (Bowling et al., 2018; Chiandetti et al., 2011). Especially important may be the prevalence of nonlinear phenomena in a tested species’ vocalizations as they can occlude harmonic structure. The review also describes how synchronous vocalizing may have shaped human consonance preferences and how differing vocal ranges between sexes may facilitate the emergence of consonance preference.

The results from McPherson et al. (2020) suggest that consonance may be less about pleasantness per se than about sounds merging to sound “as one.” Operant research in nonhuman animals could test whether this perception is

the same across species. Such research could train animals to discriminate between single tones played back and two tones played back simultaneously. Our hypothesis would suggest that animals with clear harmonic vocal output should show most perceptual merging of simultaneously played octaves, then perfect fifths, then perfect fourths, and so on, exactly matching the harmonic series in natural vocalizations.

When considering experimental approaches to assessing consonance preference, place preference paradigms such as Chiandetti and Vallortigara (2011) appear promising, as they are relatively easy and cheap to implement and can easily be adjusted for maximum ecological relevance. Furthermore, the production of consonance by animals appears to be understudied. When analyzing species' vocalizations, researchers should be attentive for consonant tone intervals occurring either in sequence or in synchrony. Production could be studied more directly as well. We hypothesized here that a vocal learning species may benefit from a perceptual frame favoring harmonic sounds so as to facilitate imitating sounds from conspecifics over (mostly nonharmonic) environmental sounds. A paradigm testing whether a vocal mimicking species would indeed preferentially mimic harmonic sounds or consonant tone sequences was suggested in Wagner et al. (2020). In such a paradigm, animals could be reared in isolation from tutors and be tutored by playback sounds, some of which could be manipulated to be less harmonic.

Regarding the choice of test species, highly harmonic vocal output and synchronized vocalization are perhaps more important than vocal learning and sexually dimorphic vocal ranges. Gibbons (*Hylobatidae*) have been shown to vocalize in duets as well as in choruses (e.g., Clarke et al., 2006) using harmonic sounds. In addition to being primates, and thereby relatively closely related to humans, they appear as highly promising subjects to test for consonance preference as well as for general musicality research. It would be especially interesting to study how juveniles interact in chorusing and duetting. Finally, as coordinated vocalization appears to be distributed unevenly among clades (Ravignani et al., 2014), it makes sense to consider animals that are more distantly related to humans, such as insects (Cator et al., 2009).

On an additional note: Research has suggested that early hominid chorusing may have been used as a deterrent to predators (Jordania, 2006), potentially similar to how gibbons use song as a defence against predators (Clarke et al., 2006). Whether consonant sounds have a stronger deterrent effect on large predators could be tested empirically, perhaps simply with a place preference paradigm

where tested predator species are free to spend time with either a sound source playing consonant or dissonant sounds (e.g., Wagner et al., 2020).

Timbre

Saliency of timbre appears to be especially constrained by having different vocal ranges and vocalizing together. As timbre has been perhaps the musicality aspect studied least in cross-species comparisons, the way that different species perceive timbre remains unclear. Bregman et al.'s (2016) paradigm, which showed that European starlings primarily depend on spectral shape in pitch tasks, could be employed with more species to determine whether the separation of pitch and timbre is humanlike. Additionally, different species' ability of parsing timbre could be compared with that of humans using the paradigm from Hoeschele et al. (2014). This comparison would allow additional insight into how timbre relates to the four traits we discussed here.

A distinction sometimes made between vocal learning (the ability to learn sounds from conspecifics) and vocal mimicry (the ability to also learn sounds from heterospecifics or the environment; Kelley et al., 2008) could point to the relevance of timbre in other species. Just like possessors of absolute pitch identify the key of a piece more quickly if it is played on their primary instrument (Marvin & Brinkman, 2000), many animals may have difficulty parsing constituent frequencies of complex sounds with which they are not familiar. However, they might be able to learn novel sounds if the sounds are presented in the correct timbre (i.e., if it sounds like a conspecific). Some bird species have been shown to learn songs of different species but preferentially copy conspecifics if they are given a choice (see Wooley, 2012, for a review). Some bird species also do better in pitch-related experiments when their own vocalizations instead of pure tones or other sounds were used to create stimuli (see Hoeschele, 2017, for a review). As such, the distinction between vocal learning and vocal mimicry may be continuous and influenced by familiarity, perhaps on a similar gradient to that of vocal learning and intermediary abilities (Wirthlin et al., 2019) which would further support a modular perspective on vocal learning.

The Connection to Social Bonding

Several hypotheses propose how and why human musical behavior may be rooted in biology. A recent review by Savage et al. (2020) presents a compelling synthesis of a large proportion of these hypotheses in the "music and social bonding hypothesis" by presenting

music as a “coevolved system for social bonding” (p. XX). We are partial to this hypothesis, as human music indeed seems to occur in contexts that relate to social bonding, be it many individuals singing together in choirs, in churches, at campfires, or in pop music concerts, or pairs, such as a parent singing a lullaby to a child or a couple dancing. Singing and moving to music seem to bring humans together. Exactly how music facilitates social bonding remains unclear, but synchronization may play a crucial role. Synchronization can facilitate social bonding (Cross et al., 2016; Feng et al., 2020; Hu et al., 2017; Kokal et al., 2011), perhaps resulting in a perceptual/motor sensation of “togetherness” or “sameness” based on a common neural encoding referred to as *self–other overlap* in psychology (e.g., Feng et al., 2020). The way that music can facilitate synchronization is perhaps most obvious with the musical property of rhythm: The perception of the beat in music is what allows humans to time their musical playing and dancing together as opposed to simply making music or moving to music within the same time period. Two humans who follow the same beat automatically act synchronized. However, abilities that allow us to intonate the same melodies together and to create harmonies, blending our voices to sound “as one,” may be just as central to musical synchronization as the ability to move to synchronize to a musical beat. Yet the link between synchrony, bonding, and spectral features of sound including pitch and timbre is less clear and has not been broadly considered. The aim of this section is to make this possible connection of spectral features and social bonding explicit with regards to octave equivalence, consonance, and timbre.

Singing together and harmonizing using the octave results in maximum possible harmonic overlap if the singers cannot produce the same fundamental frequency (as constrained by different vocal ranges). This leads to a perceptual fusion of the two sung notes (McPherson et al., 2020), which is stronger for consonant and less strong for dissonant intervals. Notice how this interaction is constrained by having clear vocal harmonics, by simultaneous vocalization, by having different vocal ranges, and by the ability of vocal learning (e.g., of new songs). The created perceptual fusion extends beyond octave equivalence and consonance into timbre. When we sing together, our voices perceptually merge to become one, not only in the perception of harmonic information but also in vocal quality. The timbre (i.e., our individual voice quality) dissolves into a sound not quite like any of the individual singing voices. This is especially evident when using harmony as just described, because the harmonic information of one voice overlaps the other voice. It is easy to see how perceiving

this perceptual fusion with another voice could facilitate self–other overlap.

Ancestrally, such bonding by self–other overlap could have been men and women in mating rituals or parents and their children with the side effect of training vocal communication and imitation. It is, however, also conceivable that our ancestors used these relations (i.e., the manipulation of the perception of harmony and timbre) to defend themselves against other groups or predators (Jordania, 2006) in a display of group strength, as suggested by Mehr et al. (2020). (Note, however, that Mehr et al. are not partial to the music and social bonding hypothesis.) Yet using musical activity as a display of group strength may not necessarily be at odds with its engendering social bonding at the same time (see Bowling et al., 2021). The idea that vocal activity could have been used as a defense is made more plausible by the fact that gibbons similarly use chorusing as a defence mechanism (Clarke et al., 2006). Using harmony, a human chorusing group could, for example, create the impression that instead of a group of small vulnerable animals they were one unified larger beast, thereby also making it difficult to acoustically detect a single individual in the group.

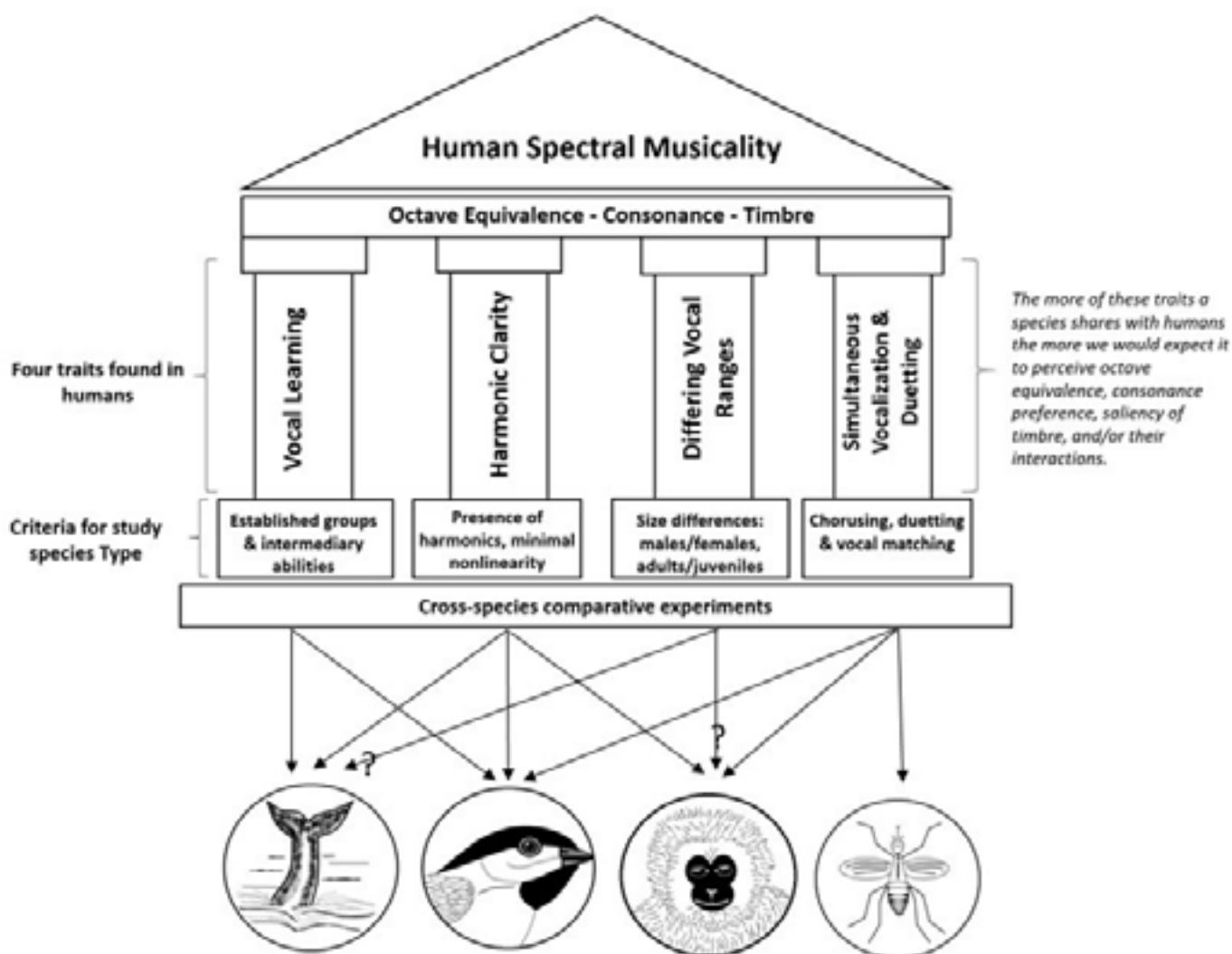
However, theories of ancestral musical activity remain partly speculative, and more research is needed. So far, research has shown that singing together can indeed facilitate social bonding; for example, Pearce et al. (2015) showed that groups that sang together bonded faster than groups practicing other activities. However, studies such as Cross et al. (2016) showed that rhythmic synchronization also facilitates social bonding. As such, the effect observed by a study such as Pearce et al. could also be due to rhythmic synchronization (i.e., synchronization on the vertical plane of sound) without necessarily being facilitated by synchronization in the harmonic spectrum (the horizontal plane of sound), as singing together entails both types of synchronization. To test whether social bonding can be facilitated by synchronization in the harmonic spectrum alone—perhaps by self–other overlap mediated by perceptual fusion—study designs where rhythmic matching and harmonic matching are separated appear a promising endeavor. Possible designs could ask subjects to match harmonic information of single notes, which are not embedded in a rhythmical structure and evaluate bonding by similar parameters as Cross et al. or Pearce et al.

Research that entails matching harmonic information in the vocalizations of nonhuman animals in real time would be highly interesting but is perhaps not easily implemented. Consideration of antiphonal call matching emerges as a promising path, as such call matching can

be understood as nonsimultaneous matching of harmonic information. Call matching in nonhuman animals is often associated with social bonding in a number of mammalian species (see King & McGregor, 2016, for a review). In the form of call-and-response, call matching is also a common feature of much human music and has been proposed to be ancestral (Jordania, 2006). This idea that is supported by nonhuman primates having been documented to engage in call matching as well (Snowdon & Elowson, 1999). Antiphonal vocal matching could therefore be highly relevant to studying musicality, as it may open numerous comparative

avenues in species where simultaneous vocalization does not occur. However, rigorously study has not determined whether matching harmonic information at separate times, such as in antiphonal calling, can facilitate social bonding in humans. As such, it would be particularly interesting to discover (a) whether nonsimultaneous matching of harmonic information can engender self–other overlap in humans and (b) whether this matching facilitates social bonding; and beyond that, (c) whether the socially related aspects of call matching in nonhuman animals are facilitated by self–other overlap and (d) if yes, whether

Figure 2. Summary of the proposed hypotheses and future research. Vocal learning, harmonic clarity, differing vocal ranges, and simultaneous vocalization & duetting constrain octave equivalence, preference of consonance, and saliency of timbre. The more of these traits a species shares with humans, the more we expect it to share aspects of human spectral musicality. The way to test this hypothesis is through cross-species comparative experiments. As we are not aware of a species that shares all four traits with humans, we suggest testing multiple different species that share different combinations of these traits. For example, gibbons and some cetacean species have harmonic vocal output, and because of size differences between adults and juveniles, perhaps at least some species have different vocal ranges for individuals. Yet only cetaceans are considered vocal learners (Tyack, 2008), whereas gibbons are known to engage in chorusing behavior (Clarke et al., 2006). Species that share some of the four traits but have mechanisms of sound production different from mammals (e.g., the songbird syrinx or insects producing sound by stridulation) can help account for potential aspects of phylogeny and for the universality of the constraint of the four traits on spectral musicality.



this is dependent on acoustic behavior or can be elicited by matching other behaviors, such as motor synchrony. As such, the existence of effects of self–other overlap in nonhuman species and their potential impact on social behavior could be studied comparatively. This endeavor would help us understand whether self–other overlap is indeed a mechanism that can facilitate social behavior in nonhuman species and whether there are constraints that favor self–other overlap by acoustic behavior over other perceptual domains.

Conclusion

We have suggested in this review that spectral aspects of musicality such as octave equivalence, consonance preference, and salience of timbre are constrained primarily by four traits—vocal learning, harmonic clarity in vocalizations, differing vocal ranges, and simultaneous vocalization and duetting. We have shown how these four traits constrain perception of octave equivalence, consonance and dissonance, and the separation of pitch and timbre. We have also discussed how octave equivalence and salience of consonance and timbre interact to create the perception of multiple voices merging into one when in harmony. Finally, we have proposed that, as with rhythm, synchronizing harmonic information may facilitate social bonding, tying this review in with the larger idea that human musicality evolved to facilitate social bonding (Savage et al., 2020). With all our ideas and hypotheses we have proposed possible avenues for future research, summarized in Figure 2.

Studies designed to gain further understanding of the proposed connection between the four traits and spectral musicality are already in preparation by our group. However, as research in this field is still few and far between, any cross-species musicality research focusing on the spectral qualities of sound can help shed light on the origin and mechanisms of spectral musicality. As such, we actively invite other researchers to strengthen our efforts to study these four traits and their connection to musicality across a wide variety of species. Doing so could ultimately strengthen our understanding of the origin of music in our species.

References

- Akre, K. L., Bernal, X., Rand, A. S., & Ryan, M. J. (2014). Harmonic calls and indifferent females: no preference for human consonance in an anuran. *Proceedings of the Royal Society B*, *281*(1789), 20140986. <https://doi.org/10.1098/rspb.2014.0986>
- Allen, O. (1967). Octave discriminability of musical and non-musical subjects. *Psychonomic Science*, *7*(12), 421–422. <https://doi.org/10.3758/BF03331154>
- Arnal, L., Flinker, A., Kleinschmidt, A., Giraud, A., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology*, *25*(15), 2051–2056. <https://doi.org/10.1016/j.cub.2015.06.043>
- Athanasopoulos, G., Eerola, T., Lahdelma, I., & Kaliakatos-Papakostas, M. (2021). Harmonic organisation conveys both universal and culturespecific cues for emotional expression in music. *PLoS ONE*, *16*(1), e0244964. <https://doi.org/10.1371/journal.pone.0244964>
- Berg, M., Fuchs, M., Wirkner, K., Loeffler, M., Engel, C., & Berger, T. (2017). The speaking voice in the general population: Normative data and associations to socio-demographic and lifestyle factors. *Journal of Voice*, *31*(2), 257.e13–257.e24. <https://doi.org/10.1016/j.jvoice.2016.06.001>
- Blackwell, H. R., & Schlosberg, H. (1943). Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *Journal of Experimental Psychology*, *33*(5), 407–419. <https://doi.org/10.1037/h0057863>
- Bowling, D. L., Hoeschele, M., & Dunn, J. D. (2021). Progress without exclusion in defining a framework for the evolution of music. *Behavioral and Brain Sciences*, *44*, Article e97. <https://doi.org/10.1017/S0140525X-20001466>
- Bowling, D. L., Hoeschele, M., Gill, K. Z., & Fitch, W. T. (2017). The nature and nurture of musical consonance. *Music Perception*, *35*(1), 118–121. <https://doi.org/10.1525/mp.2017.35.1.118>
- Bowling, D. L., & Purves, D. (2015). A biological rationale for musical consonance. *Proceedings of the National Academy of Sciences*, *112*(36), 11155–11160. <https://doi.org/10.1073/pnas.1505768112>

- Bowling, D. L., Purves, D., & Gill, K. Z. (2018). Vocal similarity predicts the relative attraction of musical chords. *Proceedings of the National Academy of Sciences*, *115*(1), 216–221. <https://doi.org/10.1073/pnas.1713206115>
- Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proceedings of the National Academy of Sciences*, *113*(6), 1666–1671. <https://doi.org/10.1073/pnas.1515380113>
- Brooks, D. I., & Cook R. G. (2010). Chord discrimination by pigeons. *Music Perception*, *27*(3), 183–196. <https://doi.org/10.1525/mp.2010.27.3.183>
- Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, *41*(2), 229–248. <https://doi.org/10.1177/0305735611425896>
- Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *Psychology of music* (2nd ed., pp. 215–264). Academic Press. <https://doi.org/10.1016/B978-012213564-4/50008-1>
- Cator, L. J., Arthur, B. J., Harrington, L. C., & Hoy, R. R. (2009). Harmonic convergence in the love songs of the dengue vector mosquito. *Science*, *323*(5917), 1077–1079. <https://doi.org/10.1126/science.1166541>
- Celma-Miralles, A., & Toro, J. M. (2018). Beat perception in a non-vocal learner: Rats can identify isochronous beats. In C. Cuskley, M. Flaherty, H. Little, L. McCrohon, A. Ravignani, & T. Verhoef (Eds.), *Proceedings of the 12th International Conference on the Evolution of Language* (pp. 76–79). <https://doi.org/10.12775/3991-1.015>
- Chen, J. C. Y. (1996). *Early Chinese work in natural science: A re-examination of the physics of motion, acoustics, astronomy and scientific thought*. Hong Kong University Press.
- Chiandetti, C., & Vallortigara, G. (2011). Chicks like consonant music. *Psychological Science*, *22*(10), 1270–1273. <https://doi.org/10.1177%2F0956797611418244>
- Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS ONE*, *1*(1), e73. <https://doi.org/10.1371/journal.pone.0000073>
- Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (*Zalophus californianus*) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology*, *127*(4), 412–442. <https://doi.org/10.1037/a0032345>
- Cousineau, M., McDermott, J., & Peretz, I. (2012). The basis of musical consonance as revealed by congenital amusia. *Proceedings of the National Academy of Sciences*, *109*(48), 19858–19863. <https://doi.org/10.1073/pnas.1207989109>
- Crespo-Bojorque, P., & Toro, J. M. (2015). The use of interval ratios in consonance perception by rats (*Rattus norvegicus*) and humans (*Homo sapiens*). *Journal of Comparative Psychology*, *129*(1), 42–51. <https://doi.org/10.1037%2Fa0037991>
- Crocker, R. L. (1963). Pythagorean mathematics and music. *The Journal of Aesthetics and Art Criticism*, *22*(3), 189–198.
- Cross, L., Wilson, A. D., & Golonka, S. (2016). How moving together brings us together: When coordinated rhythmic movement affects cooperation. *Frontiers in Psychology*, *7*, 1983. <https://doi.org/10.3389/fpsyg.2016.01983>
- Cynx, J. (1993). Auditory frequency generalization and a failure to find octave generalization in a songbird, the European starling (*Sturnus vulgaris*). *Journal of Comparative Psychology*, *107*(2), 140–146. <https://doi.org/10.1037/0735-7036.107.2.140>
- Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America*, *76*(1), 57–66. <https://doi.org/10.1121/1.391006>
- Doolittle, E., & Brumm, H. (2012). O Canto do Uirapuru: Consonant intervals and patterns in the song of the musician wren. *Journal of Interdisciplinary Music Studies*, *6*(1), 55–85. <https://doi.org/10.4407/jims.2013.10.003>
- Doolittle, E. L., Gingras, B., Endres, D. M., & Fitch, W. T. (2014). Overtone-based pitch selection in hermit thrush song: In expected convergence with scale construction in human music. *Proceedings of the National Academy of Sciences*, *111*(46), 16616–16621. <https://doi.org/10.1073/pnas.1406023111>

- Fannin, H. A., & Braud, W. G. (1971). Preference for consonant over dissonant tones in the albino rat. *Perceptual and Motor Skills*, 32(1), 191–193. <https://doi.org/10.2466/pms.1971.32.1.191>
- Feng, X., Sun, B., Chen, C., Li, W., Wang, Y., Zhang, W., Xiao, W., & Shao, Y. (2020). Self–other overlap and interpersonal neural synchronization serially mediate the effect of behavioral synchronization on prosociality. *Social Cognitive and Affective Neuroscience*, 15(2), 203–214. <https://doi.org/10.1093/scan/nsaa017>
- Fitch, W. T., Neubauer, J., & Herzog, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, 63(3), 407–418. <https://doi.org/10.1006/anbe.2001.1912>
- Hacki, T. (1999). Tonhöhen- und Intensitätsbefunde bei Stimmgeübten Vergleichende Sprechstimmfeld-, Rufstimmfeld- und Singstimmfeldmessung [Findings of pitch and intensity in practiced voices: Comparative measurement of speaking voice field, calling voice field and singing voice field]. *HNO*, 47, 809–815. <https://doi.org/10.1007/s001060050464>
- Hammerschmidt, K., Newman, J. D., Champoux, M., & Suomi, S. J. (2000). Changes in rhesus macaque “coo” vocalizations during early development. *Ethology*, 106(10), 873–886. <https://doi.org/10.1046/j.1439-0310.2000.00611.x>
- Harma, A., & Somervuo, P. (2004). Classification of the harmonic structure in bird vocalization. *IEEE International Conference on Acoustics, Speech, and Signal Processing*. <https://doi.org/10.1109/ICASSP.2004.1327207>
- Hattori, Y., & Tomonaga M. (2019). Rhythmic swaying induced by sound in chimpanzees (*Pan troglodytes*). *Proceedings of the National Academy of Sciences*, 117(2), 936–942. <https://doi.org/10.1073/pnas.1910318116>
- Hoeschele, M. (2017). Animal pitch perception: Melodies and harmonies. *Comparative Cognition & Behavior Reviews*, 12, 5–18. <https://doi.org/10.3819/CCBR.2017.120002>
- Hoeschele, M., & Bowling, D. (2016). Sex differences in rhythmic preferences in the budgerigar (*Melopsittacus undulatus*): A comparative study with humans. *Frontiers in Psychology*, 7(1543), 1–10. <https://doi.org/10.3389/fpsyg.2016.01543>
- Hoeschele, M., Cook, R. G., Guillette, L. M., Hahn, A. H., & Sturdy, C. B. (2014). Timbre influences chord discrimination in black-capped chickadees (*Poecile atricapillus*) but not humans (*Homo Sapiens*). *Journal of Comparative Psychology*, 128(4), 387–401. <https://doi.org/10.1037/a0037159>
- Hoeschele, M., Merchant, H., Kikuchi, Y., & Ten Cate, C. (2015). Searching for the origins of musicality across species. *Philosophical Transactions of the Royal Society B*, 370(1664), 20140094. <https://doi.org/10.1098/rstb.2014.0094>
- Hoeschele, M., Weisman, R. G., & Sturdy, C. B. (2012). Pitch chroma discrimination, generalization and transfer tests of octave equivalence in humans. *Attention Perception & Psychophysics*, 74(8), 1742–1760. <https://doi.org/10.3758/s13414-012-0364-2>
- Hoeschele, M., Weisman, R. G., Sturdy, C. B., Hahn, A., & Guillette, L. (2013). Chickadees fail standardized operant tests for octave equivalence. *Animal Cognition*, 16(4), 599–609. <https://doi.org/10.1007/s10071-013-0597-z>
- Honing, H., Merchant, H., Háden, G., Prado, L., & Bartolo, R. (2012). Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat. *PLoS ONE*, 7, e51369. <https://doi.org/10.1371/journal.pone.0051369>
- Honing, H., Bouwer, F. L., Prado, L., & Merchant, H. (2018). Rhesus monkeys (*Macaca mulatta*) sense isochrony in rhythm, but not the beat: Additional support for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, 12, 475. <https://doi.org/10.3389/fnins.2018.00475>
- Hove, M. J., Sutherland, M. E., & Krumhansl, C. L. (2009). Ethnicity effects in relative pitch. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, 31, 2256–2261.
- Hu, Y., Hu, Y., Li, X., Pan, Y., & Cheng, X. (2017). Brain-to-brain synchronization across two persons predicts mutual prosociality. *Social Cognitive and Affective Neuroscience*, 12(12), 1835–1844. <https://doi.org/10.1093/scan/nsx118>

- Hulse, S. H., Bernard, D. J., & Braaten, R. F. (1995). Auditory discrimination of chord-based spectral structures by European starlings (*Sturnus vulgaris*). *Journal of Experimental Psychology: General*, 124(4), 409–423. <https://doi.org/10.1037/0096-3445.124.4.409>
- Izumi, A. (2000). Japanese monkeys perceive sensory consonance of chords. *Journal of the Acoustical Society of America*, 108(6), 3073–3078. <https://doi.org/10.1121/1.1323461>
- Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, 29(19), 3229–3243. <https://doi.org/10.1016/j.cub.2019.08.020>
- Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, 366(6461), 50–54. <https://doi.org/10.1126/science.aax0287>
- Jordania, J. (2006). *Who asked the first question?* Logos.
- Kallman, H. J. (1982). Octave equivalence as measured by similarity ratings. *Perception & Psychophysics*, 32(1), 37–49. <https://doi.org/10.3758/BF03204867>
- Kelley, L. A., Coe, R. L., Madden, R., & Healy, S. D. (2008). Vocal mimicry in songbirds. *Animal Behaviour*, 76(3), 521–528. <https://doi.org/10.1016/j.anbehav.2008.04.012>
- Kent, R. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech, Language, and Hearing Research*, 19(3), 421–447. <https://doi.org/10.1044/jshr.1903.421>
- King, S. L., & McGregor, P. K. (2016). Vocal matching: The what, the why and the how. *Biology Letters*, 12(10), 20160666. <https://doi.org/10.1098/rsbl.2016.0666>
- Klapuri, A., & Davy, M. (2006). *Signal processing methods for music transcription*. Springer. <https://doi.org/10.1007/0-387-32845-9>
- Koda, H., Basile, M., Olivier, M., Remeuf, K., Nagumo, S., & Blois-Heulin, L.A. (2013). Validation of an auditory sensory reinforcement paradigm: Campbell's monkeys (*Cercopithecus campbelli*) do not prefer consonant over dissonant sounds. *Journal of Comparative Psychology*, 127(3), 265–271. <https://doi.org/10.1037/a0031237>
- Kokal, I., Engel, A., Kirschner, S., & Keysers, C. (2011). Synchronized drumming enhances activity in the caudate and facilitates prosocial commitment: If the rhythm comes easily. *PLoS ONE*, 6(11), e27272. <https://doi.org/10.1371/journal.pone.0027272>
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford University Press.
- Langner, G., & Ochse, M. (2006). The neural basis of pitch and harmony in the auditory system. *Musicae Scientiae*, 10(1), 185–208. <https://doi.org/10.1177/102986490601000109>
- Lavenex, P. B. (1999). Vocal production mechanisms in the budgerigar (*Melopsittacus undulatus*): The presence and implications of amplitude modulation. *Journal of the Acoustical Society of America*, 106(1), 491–505. <https://doi.org/10.1121/1.427079>
- Martins, R., Trancoso, I., Abad, A., & Meinedo H. (2009). Detection of children's voices. *Proceedings of the Iberian SLTech 2009*, 77–80.
- Marvin, E. W., & Brinkman, A. R. (2000). The effect of key color and timbre on absolute pitch recognition in musical contexts. *Music Perception*, 18(2), 111–137. <https://doi.org/10.2307/40285905>
- Masataka, N. (2006). Preference for consonance over dissonance by hearing newborns of deaf parents and of hearing parents. *Developmental Science*, 9(1), 46–50. <https://doi.org/10.1111/j.1467-7687.2005.00462.x>
- Matrosova, V., Volodin, I., Volodina, E., & Babitsky, A. (2007). Pups crying bass: Vocal adaptation for avoidance of age-dependent predation risk in ground squirrels? *Behavioral Ecology and Sociobiology*, 62(2), 181–191. <https://doi.org/10.1007/s00265-007-0452-9>
- McDermott, J. H., & Hauser, M. (2004). Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition*, 94(2), B11–B21. <https://doi.org/10.1016/j.cognition.2004.04.004>
- McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, 20(11), 1035–1041. <https://doi.org/10.1016/j.cub.2010.04.019>

- McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, *535*(7613), 547–550. <https://doi.org/10.1038/nature18635>
- McPherson, M. J., Dolan, S. E., Durango, A., Ossandon, T., Valdés J., Undurraga, E. A., Jacoby, N., Godony, R. A., & McDermott, J. H. (2020). Perceptual fusion of musical notes by native Amazonians suggests universal representations of musical intervals. *Nature Communications*, *11*(1), Article 2786. <https://doi.org/10.1038/s41467-020-16448-6>
- Mehr, S. A., Krasnow, M., Bryant, G. A., & Hagen, E. H. (2020). Origins of music in credible signaling. *Behavioral and Brain Sciences*, *44*. <https://doi.org/10.31234/osf.io/nrqb3>
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 647–681. <https://doi.org/10.1038/nn1082>
- Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, *24*(1), 99–104. <https://doi.org/10.1525/mp.2006.24.1.99>
- Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, *19*(10), 827–830. <https://doi.org/10.1016/j.cub.2009.03.038>
- Pearce, E., Launay, J., & Dunbar, R. I. (2015). The ice-breaker effect: Singing mediates fast social bonding. *Royal Society Open Science*, *2*(10), Article 150221. <https://doi.org/10.1098/rsos.150221>
- Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreoli, G., Rovelli, R., & Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences*, *107*(10), 4758–4763. <https://doi.org/10.1073/pnas.0909074107>
- Peter, B., Foster, B., Haas, H., Middleton, K., & McKibben, K. (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/j.jvoice.2014.06.011>
- Peter, B., Larkin, T., & Stoel-Gammon, C. (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. <https://doi.org/10.1121/1.3203993>
- Peter, B., Stoel-Gammon, C., & Kim, D. (2008). Octave equivalence as an aspect of stimulus-response similarity during nonword and sentence imitations in young children. *Proceedings of the 4th International Conference on Speech Prosody, SP 2008*, 731–734.
- Pemberton, C., McCormack, P., & Russell, A. (1998). Have women's voices lowered across time? A cross sectional study of Australian women's voices. *Journal of Voice*, *12*(2), 208–213. [https://doi.org/10.1016/S0892-1997\(98\)80040-4](https://doi.org/10.1016/S0892-1997(98)80040-4)
- Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 976–986. <https://doi.org/10.1037/0096-1523.20.5.976>
- Platinga, J., & Trehub, S. E. (2014). Revisiting the innate preference for consonance. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 40–49. <https://doi.org/10.1037/a0033471>
- Porter, D., & Neuringer, A. (1984) Music discrimination by pigeons. *Journal of Experimental Psychology, Animal Behaviour Processes*, *10*(2), 138–148. <https://doi.org/10.1037/0097-7403.10.2.138>
- Puts, D. A., Hill, A. K., Bailey, D. H., Walker, R. S., Rendall, D., Wheatley, J. R., Welling, L. L., Dawood, K., Cárdenas, R., Burriss, R. P., Jablonski, N. G., Shriver, M. D., Weiss, D., Lameira, A. R., Apicella, C. L., Owren, M. J., Barelli, C., Glenn, M. E., & Ramos-Fernandez, G. (2016). Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proceedings of the Royal Society B*, *283*(1829), Article 20152830. <https://doi.org/10.1098/rspb.2015.2830>
- Ravignani, A., Bowling, D. L., & Fitch W. T. (2014). Chorusing, synchrony, and the evolutionary functions of rhythm. *Frontiers in Psychology*, *5*, Article 1118. <https://doi.org/10.3389/fpsyg.2014.01118>
- Reichmuth, C., & Casey, C. (2014). Vocal learning in seals, sea lions, and walruses. *Current Opinion in Neurobiology*, *28*, 66–71. <https://doi.org/10.1016/j.conb.2014.06.011>

- Richards, D. G., Wolz, J. P., & Herman, L. M. (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. <https://doi.org/10.1037/0735-7036.98.1.10>
- Richner, H. (2016). Interval singing links to phenotypic quality in a songbird. *Proceedings of the National Academy of Sciences*, 113(45), 12763–12767. <https://doi.org/10.1073/pnas.1610062113>
- Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2020). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, 44. <https://doi.org/10.1017/S0140525X20000333>
- Sclan, J. (1999). The function and significance of inter-species acoustic cues in the transformation of budgerigar (*Melopsittacus undulatus*) sounds into “speech.” *International Journal of Comparative Psychology*, 12(3), 111–152.
- Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, 19(10), 831–836. <https://doi.org/10.1016/j.cub.2009.03.061>
- Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infant listeners. *Psychological Science*, 7(5), 272–277. <https://doi.org/10.1111%2Fj.1467-9280.1996.tb00373.x>
- Schwartz, D. A., Howe, C. Q., & Purves, D. (2003). The statistical structure of human speech sounds predicts musical universals. *Journal of Neuroscience*, 23(18), 7160–7168. <https://doi.org/10.1523/JNEUROSCI.23-18-07160.2003>
- Snowdon, C. T., & Elowson, A. M. (1999). Pygmy marmosets modify call structure when paired. *Ethology*, 105(10), 893–908. <https://doi.org/10.1046/j.1439-0310.1999.00483.x>
- Snowdon, C. T., & Teie, D. (2010). Affective responses in tamarins elicited by species-specific music. *Biology Letters*, 6(1), 30–32. <https://doi.org/10.1098/rsbl.2009.0593>
- Sugimoto, T., Kobayashi, H., Nobuyoshi, N., Kiriya, Y., Takeshita, H., Nakamura, T., & Hashiya, K. (2010). Preference for consonant music over dissonant music by an infant chimpanzee. *Primates*, 51(1), 7–12. <https://doi.org/10.1007/s10329-009-0160-3>
- Sugiura, H. (1998). Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Animal Behaviour*, 55(3), 673–687. <https://doi.org/10.1006/anbe.1997.0602>
- Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music Perception*, 1(3), 276–295. <https://doi.org/10.2307/40285261>
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, 85(4), Article 1699. <https://doi.org/10.1121/1.397959>
- Toro, J., & Crespo, P. (2017). Consonance processing in the absence of relevant experience: Evidence from nonhuman animals. *Comparative Cognition & Behavior Reviews*, 12, 33–44. <https://doi.org/10.3819/CCBR.2017.120004>
- Trainor, L. J. & Heinmiller, B. M. (1998). The development of evaluative responses to music. *Infant Behavior and Development*, 21(1), 77–88. [https://doi.org/10.1016/S0163-6383\(98\)90055-8](https://doi.org/10.1016/S0163-6383(98)90055-8)
- Trainor, L. J., Tsang, C. D., & Cheung, V. H. M. (2002). Preference for sensory consonance in 2- and 4-month-old infants. *Music Perception*, 20(2), 187–194. <https://doi.org/10.1525/mp.2002.20.2.187>
- Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, 6(7), 669–673. <https://doi.org/10.1038/nm1084>
- Tu, H. W., Osmanski, M. S., & Dooling, R. J. (2011). Learned vocalizations in budgerigars (*Melopsittacus undulatus*): The relationship between contact calls and warble song. *Journal of the Acoustical Society of America*, 129(4), 2289–2299. <https://doi.org/10.1121/2F1.3557035>
- Tyack, P. L. (2008). Convergence of calls as animals form social bonds, active compensation for noisy communication channels, and the evolution of vocal learning in mammals. *Journal of Comparative Psychology*, 122(3), 319–331. <https://doi.org/10.1037/a0013087>

- van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38(3), 253–265. <https://doi.org/10.1177/002383099503800303>
- Virtala, P., & Tervaniemi, M. (2016). Neurocognition of major-minor and consonance-dissonance. *Music Perception*, 34(4), 387–404. <https://doi.org/10.1525/mp.2017.34.4.387>
- Wagner, B., Bowling, D. L., & Hoeschele, M. (2020). Is consonance attractive to budgerigars? No evidence from a place preference study. *Animal Cognition*, 23(5), 973–987. <https://doi.org/10.1007/s10071-020-01404-0>
- Wagner, B., & Hoeschele, M. (2022). *Budgerigar (Melopsittacus undulatus) vocal training to uncover mechanism that enables mimicry of the human voice* [Manuscript in preparation]. Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria.
- Wagner, B., Mann, D. C., Afroozeh, S., Staubmann, G., & Hoeschele, M. (2019). Octave equivalence is not linked to vocal mimicry: Budgerigars fail standardized operant tests for octave equivalence. *Behaviour*, 156(5–8), 480–504. <https://doi.org/10.1163/1568539X-00003538>
- Watanabe, S., Uozumi, M., & Tanaka, N. (2005). Discrimination of consonance and dissonance in Java sparrows. *Behavioural Processes*, 70(2), 203–208. <https://doi.org/10.1016/j.beproc.2005.06.001>
- Weirich, M., & Simpson, A. P. (2018). Gender identity is indexed and perceived in speech. *PLoS ONE*, 13(12), Article e0209226. <https://doi.org/10.1371/journal.pone.0209226>
- Wirthlin, M., Chang, E. F., Knörnschild, M., Krubitzer, L. A., Mello, C. V., Miller, C. T., Pfenning, A. R., Vernes, S. C., Tchernichovski, O., & Yartsev, M. M. (2019). A modular approach to vocal learning: disentangling the diversity of a complex behavioral trait. *Neuron*, 104(1), 87–99. <https://doi.org/10.1016/j.neuron.2019.09.036>
- Wooley, S. M. (2012). Early experience shapes vocal neural coding and perception in songbirds. *Developmental Psychobiology*, 54(6), 612–631. <https://doi.org/10.1002/dev.21014>
- Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology*, 129(3), 291–307. <https://doi.org/10.1037/0096-3445.129.3.291>
- Zamorano-Abramson, J., Hernández-Lloreda, V., García, L., Colmenares, F., Aboitiz, F., & Call, J. (2018). Imitation of novel conspecific and human speech sounds in the killer whale (*Orcinus orca*). *Proceedings of the Royal Society B*, 285(1871), 20172171. <https://doi.org/10.1098/rspb.2017.2171>
- Zentner, M. R., & Kagan, J. (1996). Perception of music by infants. *Nature*, 383(6595), Article 29. <https://doi.org/10.1038/383029a0>
- Zentner, M. R., & Kagan, J. (1998). Infants' perception of consonance and dissonance in music. *Infant Behavior and Development*, 21(3), 483–492. [https://doi.org/10.1016/S0163-6383\(98\)90021-2](https://doi.org/10.1016/S0163-6383(98)90021-2)