Complex vocal learning and three-dimensional mating environments

Reference:
Verpooten Jan.- Complex vocal learning and three-dimensional mating environments
Biology and philosophy - ISSN 0169-3867 - 36:2(2021), 12
Full text (Publisher's DOI): https://doi.org/10.1007/S10539-021-09786-2
To cite this reference: https://hdl.handle.net/10067/1766530151162165141

Complex vocal learning and three-dimensional mating environments

This is a pre-copyedit version of an article accepted in Biology & Philosophy

Jan Verpooten

KU Leuven, Leuven, Belgium

University of Antwerp, Antwerpen, Belgium

jan.verpooten@kuleuven.be

ORCID: 0000-0001-5547-7330

Abstract

Complex vocal learning, the capacity to imitate new sounds, underpins the evolution of animal vocal cultures and song dialects and is a key prerequisite for human speech and song. Due to its relevance for the understanding of cultural evolution and the biology and evolution of language and music, the trait has gained much scholarly attention. However, while we have seen tremendous progress with respect to our understanding of its morphological, neurological and genetic aspects, its peculiar phylogenetic distribution has remained elusive. Intriguingly, animals as distinct as hummingbirds and humpback whales share well-developed vocal learning capacity in common with humans, while this ability is quite limited in nonhuman primates. Yet, solving this 'vocal learning conundrum' may shed light on the constraints ancestral humans overcame to unleash their vocal capacities. To this end I consider major constraints and functions that have been proposed. I highlight an especially promising ecological constraint, namely the spatial dimensionality of the environment. Based on an informal comparative review, I suggest that complex vocal learning is associated with three-dimensional habitats such as air and water. I argue that this is consistent with recent theoretical advances – i.e., the coercion-avoidance and dimensionality hypotheses – and with the long-standing hypothesis that mate choice is a major driver of the origin and evolution of complex vocal learning. However, I stress that multiple functions may apply and that quantitative phylogenetic comparative methods should be employed to finally resolve the issue.

Keywords: complex vocal learning, mate choice, coercion avoidance, dimensionality hypothesis, cross-species comparison, musical protolanguage

Intro

Complex vocal learning or the ability to imitate novel sounds from the environment is a rare and phylogenetically sparsely distributed trait that humans have in common with animals as distinct as hummingbirds and humpback whales, but not with any nonhuman primate (Fitch, 2017; Janik & Slater, 1997; Searcy, 2019; Tyack, 2020). It is an essential prerequisite for speech and song (Arriaga & Jarvis, 2013; Belyk & Brown, 2017), and therefore has important implications for the understanding of brain mechanisms and disorders of spoken language (Jarvis, 2019). It underpins the evolution of animal vocal cultures and song dialects, as such highlighting the importance of cultural evolution in nonhuman animals as well as providing useful points of comparison with human cultural evolution (Killin, 2016; Laland & Janik, 2006). Its relevance to the understanding of the nature and evolution of music, language and culture is thus substantial (cf. Mehr et al., 2020; Savage et al. 2020).

Given its pivotal role and broad relevance, it has drawn considerable cross-disciplinary research effort, which has substantially increased the understanding of its functional, developmental, genetic (e.g., FOXP2) and morphological underpinnings (vocal and auditory apparatus) across species (Lattenkamp & Vernes, 2018). However, despite discussions of potential advantages and constraints on its evolution (Janik & Slater, 1997; Jarvis, 2006; Jarvis, 2019; Nottebohm, 1972; Nowicki & Searcy, 2014), it remains unclear why well-developed vocal learning is limited to a few bird and mammal groups (including humans). This paper aims to tackle this 'vocal learning conundrum' (Larsson & Abbott, 2018) and shed light on the origins of complex vocal learning in the human lineage.

After introducing important distinctions and the taxonomic distribution of complex vocal learning, I review the major functions and constraints that have been proposed, paying special attention to the spatial dimensionality of the physical environment. This is a particularly promising ecological constraint, initially proposed by Janik and Slater (1997), but which has never been fully explored. The idea is that in three-dimensional space (such as air or water), it is more difficult for males to monopolize females and that therefore males evolve vocal learning ability in order to produce complex courtship songs to attract females instead. I expand on this basic idea based on the female mate choice hypothesis of vocal learning (Janik & Slater, 1997; Jarvis, 2006; Jarvis, 2009; Nottebohm, 1972; Nowicki & Searcy, 2014; Okanoya, 2002), which I extend with empirical and theoretical work on male/mutual mate choice (Courtiol et al., 2016; Riebel et al., 2019; Stewart-Williams & Thomas, 2013). I also expand on it based on recent theoretical advances, namely the dimensionality hypothesis (Puts, 2010) and the coercion avoidance hypothesis (Pradhan & van Schaik, 2009). Furthermore, I back it up with an informal comparative analysis based on up-to-date findings which suggests an association between complex vocal learning

and three-dimensionality for most taxa. Specifically, it appears that over 99% of complex vocal learning species and, when taking into account phylogenetic relationships, 6 out of 8 complex vocal learning taxa inhabit essentially three-dimensional environments. Moreover, this association seems to a substantial extent mediated by (mutual) mate choice. Finally, I comparatively evaluate the explanatory potential for the origins of complex vocal learning in the human and other lineages of this 3D/mate choice hypothesis and alternative hypotheses (e.g., social cohesion and information sharing with kin: Fitch, 2005; Sewall et al., 2016) and make suggestions for further research.

Defining distinctions

Janik and Slater (2000) distinguish between vocal production learning on the one hand, which involves changing the acoustic parameters of vocalizations, and contextual learning on the other. Contextual learning includes usage learning, which involves producing an existing signal in a new context, and comprehension learning, in which receivers learn to modify their response to vocalizations produced by others. Because contextual learning has a wider taxonomic occurrence (possibly because it requires a less considerable neuronal apparatus), its evolution seems less puzzling (Nowicki & Searcy, 2014). Here I am only concerned with *vocal production learning*, which I will refer to as 'vocal learning'.

Furthermore, 'vocalization', by definition, refers only to sounds produced by the vocal organ (larynx in anurans, reptiles and mammals and the syrinx in birds) (Bradbury & Vehrencamp, 2011; Petkov & Jarvis, 2012). Thus sounds generated by other anatomical sources have commonly been excluded from discussions about vocal learning. For instance, sounds produced by teeth, tongue and lips are rather considered non-voiced or orio-facially generated sounds. This distinction make sense because the neural mechanisms and the degree of control of the different musculature for generating voiced or unvoiced sounds probably differ (Petkov & Jarvis, 2012).

A further distinction involves the level of complexity of the vocal learning behavioral phenotype and its neurological underpinnings, as captured by the stepwise continuum view on vocal learning (Arriaga & Jarvis, 2013; Jarvis, 2019). Tyack (2020) defines 'limited vocal learning' as the ability to fine-tune acoustic features of species-specific vocalizations that can develop in the absence of auditory input because innate motor programs can generate the species-specific pattern. In contrast, 'complex vocal learning' is defined by the need to hear a sound to form a learned auditory template before the animal can develop a vocalization that matches the template. This thus involves the capacity to *imitate sounds* and add them to the vocal repertoire (Janik & Slater, 1997; Tyack, 2020). While complex vocal learning depends on specialized neural pathways in the telencephalon and is taxonomically sparsely distributed, limited vocal learning is produced by central pattern generators in the brainstem and has a much broader, and less puzzling, distribution (Fitch, 2017; Tyack, 2020). Tyack (2020) also

argues that functions differ between limited and complex vocal learning and that limited vocal learning may not provide good animal models for studying complex vocal learning. Hence, here I will focus on *complex vocal learning* (somewhat misleadingly, often simply referred to as 'vocal learning' in the literature).

Taxonomic distribution of complex vocal learning

Complex vocal learning has a patchy distribution across vertebrates. Unequivocal evidence for complex vocal learning comes from experiments that verify whether an animal can learn to imitate sounds of another species or to imitate artificial sounds (Janik & Slater, 1997; Tyack, 2020). While most species use their vocal learning capacities to acquire a species-specific repertoire, some species imitate sounds from non-conspecifics as such providing indisputable evidence for the capacity of vocal imitation.

Oscine songbirds (suborder Passeri) and parrots (order Psittaciformes) have long been trained to successfully imitate speech and about 20% of songbirds are suggested to mimic the vocalizations of other species (Hindmarsh, 1984; Tyack, 2020). Hummingbirds (family Trochilidae) have been shown to replace their song when they hear new song types (Araya-Salas & Wright, 2013). Thus, in these three bird clades complex vocal learning is firmly established (Araya-Salas & Wright, 2013; Beecher & Brenowitz, 2005; Berg et al., 2012). Based on phylogenetic analysis it has been suggested that these clades represent either two or three independent origins of complex vocal learning in birds; one origin in the more distantly related hummingbirds and one in the common ancestor of parrots and oscine songbirds or two independent origins in these latter two groups (Searcy & Nowicki, 2019).

Regarding mammals, evidence suggests that vocal learning is a general, or at least highly prevalent, feature of behavior in bats (Knörnschild, 2014; Prat et al., 2015; Vernes, 2017). Knörnschild et al. (2010) report that pups of the bat *Saccopteryx bilineata* learn complex vocalization through vocal imitation of adult males, independent of their own sex and relatedness to the males, indicating complex vocal learning (but see Tyack, 2020).

Pinnipeds are semi-aquatic mammals comprising the Phocidae (e.g., harbor and grey seals), Otariidae (e.g., California sea lions and Cape fur seals), and Odobenidae (walruses). Field studies and reinforcement training experiments suggest that walruses and seals are capable of complex vocal learning. Classic evidence comes from Hoover, a captive male harbor seal that spontaneously produced a variety of human expressions including the phrase 'Hey! Hey! Come over here!' (Reichmuth & Casey, 2014). More recently, Stansbury and Janik (2019) have shown that grey seals can learn to modify emphasized frequency bands called formants to copy human vowels and melodies. However, Otariids such as sea lions do not appear to have complex vocal learning abilities (Schusterman, 2008).

Cetaceans comprise toothed wales (parvorder Odontoceti, e.g., dolphins, porpoises, belugas and sperm whales) and baleen whales (parvorder Mysticeti, e.g., humpback and blue whales). Strong evidence of complex vocal learning has been found in several species of both groups of cetaceans (Janik, 2014; Whitehead & Rendell, 2012). Imitation of the human voice has been observed in killer whales (*Orcinus orca*) and beluga whales (*Delphinapterus leucas*) (Abramson et al., 2018; Ridgway et al., 2012). Humpback whales (*Megaptera novaeangliae*) have been shown to individually track and copy changes in their song within and between populations, and bowhead whales (*Balaena mysticetus*) sing extremely diverse sets of songs with much interannual variability, providing strong evidence for complex vocal learning (Noad et al., 2000; Payne, 1984; Stafford et al., 2018).

At least two of three extant elephant species, the African bush elephant (*Loxodonta africana*) and the Asian elephant (*Elephas maximus*), are capable of complex vocal learning (Poole et al., 2005; Stoeger & Manger, 2014). An Asian elephant learned to copy the words used as commands by his trainer (Stoeger et al., 2012).

Finally, among primates, only humans are capable of complex vocal learning, as evidenced by decades of research on nonhuman primate vocalization (Fitch, 2017; Tyack, 2020). This is surprising given the importance of complex vocal learning in humans. While gelada baboons have a large repertoire of rhythmic and synchronized vocalizations (Richman, 1978, 1987), no evidence of vocal learning has been reported. And while vocal convergence (calls of individuals becoming more similar when they live together) has been observed in many nonhuman primates (e.g., marmosets: Takahashi et al., 2015; Takahashi et al., 2017), strong evidence suggests this reflects slight modification to innate vocalizations rather than vocal imitation. For instance, intensive efforts to train apes to speak have failed and disrupting learning opportunities through deafening or isolation have demonstrated little effect on normal vocalization development (Hayes & Hayes, 1952; Winter et al., 1973). Furthermore, cross-breeding experiments demonstrate genetic inheritance of variation in acoustic structure in several species (Brockelman & Schilling, 1984). This suggests that either complex vocal learning evolved independently in humans or that a primate common ancestor was capable of complex vocal learning, while the trait was subsequently lost at least eight other times. Based on parsimony, the independent evolution hypothesis seems much more likely (Petkov & Jarvis, 2012).

Hence, the phylogenetic distribution of these mammalian groups suggests five independent origins of complex vocal learning corresponding to these mammalian groups, which amounts to a total of about eight independent origins when also taking birds into account (Petkov & Jarvis, 2012; Searcy & Nowicki, 2019). The infrequency of complex vocal learning raises the question why this trait is not more widespread. Several constraints and functions of complex vocal learning have been advanced and they have been evaluated as potential explanations for this

vocal learning conundrum (Janik & Slater, 1997; Jarvis, 2006; Jarvis, 2019; Larsson & Abbott, 2018; Nottebohm, 1972; Nowicki & Searcy, 2014).

Functions

With respect to unraveling the uses for which complex vocal learning evolved, it is important to distinguish between evolutionary origins and current utility, as any trait that originally evolved for a particular function can be later coopted for another, a process that has been dubbed exaptation (Gould, 1991). Moreover, some hypothesized advantages are not relevant to the evolutionary origin of vocal learning. This is the case because, requiring complex neural adaptations, complex vocal learning is probably costly and thus would not evolve except in response to some compensating fitness benefit, and some hypothesized benefits would only accrue when the trait has already evolved (Nowicki & Searcy, 2014). Since the aim of this paper is to shed light on the independent origins of complex vocal learning across taxa in order to elucidate the vocal learning conundrum and its origin in the human lineage, I pay particular attention to whether hypothesized functions accrue to the first individuals that begin to modify their vocalizations based on the sounds they hear from conspecifics.

A first hypothesis, the *vocal dialect hypothesis*, assumes that males learn the local dialect where they are born and that local females preferentially mate with males that sing those dialects. This would thus promote assortative mating, which might benefit individuals if there are local genetic adaptations that can thereby persist (Janik & Slater, 1997; Nottebohm, 1972; Nowicki & Searcy, 2014). Taxa to which it might apply are songbirds and parrots. However, this hypothesis has been challenged on empirical and theoretical grounds (Janik & Slater, 1997; Nowicki & Searcy, 2014; but see: Payne et al., 1998; Payne et al., 2000). Furthermore, it has a hard time explaining the origin of vocal learning since small scale geographic variation does not occur until vocal learning is already in place (Nowicki & Searcy, 2014).

The *habitat matching hypothesis* takes into account that sound transmission characteristics can differ widely between habitats. If a species inhabits different habitats or if the transmission characteristics of its habitat change regularly, complex vocal learning could help to optimize transmission of signals (Hansen, 1979; Janik & Slater, 1997; Nowicki & Searcy, 2014). This hypothesis has gained empirical support in the form of evidence for acoustic adaptation of songs to habitats (e.g., Peters et al., 2012; Ríos-Chelén et al., 2012; Slabbekoorn & den Boer-Visser, 2006). Taxa to which it might apply are those that have long range vocal signals and that are philopatric to habitats such as songbirds (Nowicki & Searcy, 2014). However, Nowicki and Searcy (2014) point out that a major weakness of this hypothesis is that differences in songs between habitats is small, as indicated by a meta-analysis

(Boncoraglio & Saino, 2007), suggesting that the benefits of acoustic adaptation to the environment may be marginal.

Albeit too marginal to be a major driver, several findings support the habitat matching hypothesis in humans (Lupyan & Dale, 2016). A strong example are the whistle languages, which typically transpose vocal speech into a system of whistles. These tend to be found in places where a strong pressure to communicate over long distances is combined with difficult-to-traverse terrain. For instance, speakers of whistle language Silbo Gomero in the Canary Islands have been observed to communicate over distances of 10 km (Meyer, 2004). Conventional languages also show signs of acoustic adaptation. There is some evidence that prevalence of vowel use is higher in warmer climates because they have denser vegetation, which better propagates vowels (Maddieson et al., 2011). And languages in dryer climates are less likely to use lexical tone (i.e., the use of rising and falling pitch patterns to mark differences between words), possibly because dry air can hamper the precise vocal control required for making tonal distinctions (Everett et al., 2016).

The *social cohesion hypothesis* focuses on individual and group recognition through learned contact call production (Janik & Slater, 1997; Nowicki & Searcy, 2014). Contact calls are relatively simple vocalizations that are unique to specific individuals. Animals produce them when reuniting or coordinating behaviors (e.g., foraging) and therefore they are considered to promote social cohesion (Kondo & Watanabe, 2009; Nowicki & Searcy, 2014; Sewall, 2015). Contact calls are taxonomically widely distributed (Kondo & Watanabe, 2009), but most mammalian and bird species produce contact calls involving limited or no learning. Only some animals (e.g., elephants, dolphins and parrots) use complex vocal learning ability to produce them. For example, bottlenose dolphins have been shown to match their contact calls ('signature whistles') to those of others, probably to address and even label them, and similar uses have been observed in parrots (King & Janik, 2013; Scarl & Bradbury, 2009). An individual recognition role, Nowicki and Searcy (2014) argue, can, in principle, explain the origins of vocal learning. This is based on the reasoning that, if learning increases recognizability of an individual, this benefit would probably accrue to a learner even if no other individuals are vocal learners in its population. However, the case for an individual recognition advantage has been considered dubious, given the many examples of individual recognition based on unlearned vocalizations and since, logically, the increases in vocal repertoire size allowed by complex vocal learning should make recognition harder, not easier (Nottebohm, 1972; Nowicki & Searcy, 2014). According to Nowicki and Searcy (2014), it is more plausible that complex vocal learning enhances group recognition rather than individual recognition, particularly if animals disperse between groups so that they sometimes need to converge on a new group's signature. This is largely consistent with the idea that social

complexity (proxied by group size) drives vocal complexity (Blumstein & Armitage, 1997; Freeberg, 2006; McComb & Semple, 2005; Wilkinson, 2000), and with the related suggestion that complex socially dynamic systems with fluid social affiliations, such as fission–fusion social systems, are associated with the production of learned contact calls across diverse taxa (Sewall, 2015; Sewall et al., 2016).

Social cohesion in fluid complex social groups through group recognition has probably been a benefit of human complex vocal learning as well. Humans (hunter-gatherers) share fission-fusion social dynamics with chimpanzees and other primates (Couzin & Laidre, 2009; Marlowe, 2005; Symington, 1990), suggesting it is ancestral. Accordingly, it has been suggested that the huge diversification of human language is partially linked to the function of marking group identity (Roberts, 2010). Further, it has been suggested that singing together also fosters social closeness – even in large group contexts where individuals are not known to each other (Savage et al., 2020; Weinstein et al., 2016). However, this latter idea is contradicted by the fact that music, including elaborate song, would be needlessly costly to facilitate group social cohesion, whereas a low-cost signaling system such as language can efficiently fulfill this role (Mehr et al., 2020). In line with this criticism, Verpooten & Eens (in press) argue that, across species, social complexity and group cohesion function is associated with relatively simple vocalizations (i.e., contact calls) rather than energetically costly courtship songs. Importantly, Nowicki and Searcy (2014) assert that this group recognition hypothesis runs into the same logical problem as the vocal dialect problem – that it is unlikely that group specific signatures occur unless vocal learning has already evolved and that therefore it unlikely explains its origins (but see Sewall et al., 2016).

The *information-sharing hypothesis* focuses on the disparity in vocabulary size between modern humans, who have the ability to use an unlimitedly large class of symbols and nonhuman animals, which typically have a lexicon of a few dozen call types (Jackendoff, 1999). The hypothesis suggests that complex vocal learning evolved in the hominid lineage for vocabulary expansion, which was favored by kin selection for information sharing among relatives (Fitch, 2004, 2005; Nowicki & Searcy, 2014). Since information is shared with kin, the hypothesis entails an immediate genetic advantage for vocal learners and therefore can explain the origins of vocal learning. However, it has a quite narrow taxonomic application as it is only in humans that learned vocal signals provide environmental information to others (Nowicki & Searcy, 2014).

A final major hypothesis focuses on sexual selection and gives center stage to mate choice as driver of the evolution of complex vocal learning. This *mate choice hypothesis* builds on the fact that more complex song (e.g., with more varied syntax) and larger repertoire size are generally favored by females (Soma & Garamszegi, 2011) and that learned vocalizations may achieve greater complexity than innate vocalizations (Janik & Slater, 1997;

Jarvis, 2006; Jarvis, 2009; Nottebohm, 1972; Nowicki & Searcy, 2014; Okanoya, 2002). Therefore, a capacity for complex vocal learning enables an individual to attract mates more successfully and increase its reproductive success. Several complex vocal learning groups (oscine songbirds, hummingbirds, phocids, bats, and baleen whales) commonly use complex sounds to attract mates. Complex vocalizations may be favored by the opposite sex because they indicate mate quality (e.g., in the form of resistance against developmental stress: Schmidt et al., 2013), because they appeal to a pre-existing bias (Ryan, 1998, 2018; Ryan & Rand, 1990) or simply due to arbitrary coevolution (runaway sexual selection) (Prum, 2012, 2010, 2013).

A number of species naturally imitate heterospecific vocalizations, which is strong evidence for complex vocal learning, but raises the question why this would be preferred by the opposite sex. One hypothesis is that heterospecific vocal mimicry is simply an indirect way to enlarge the song repertoire and is driven by female preference for song complexity (Zann & Dunstan, 2008). Garamszegi et al. (2007) concluded, based on phylogenetic analysis, that heterospecific mimicry has no important function in female choice in European birds, is simply a product of 'learning mistakes' arising from the development of complex songs and has no precise adaptive role. Another hypothesis is that vocal mimicry accurately tracks fitness. For example, accuracy of vocal mimicry and the number of model species mimicked are both independently related to male mating success in satin bowerbirds and especially mimetic accuracy may provide females with important information about male quality (Coleman et al., 2007). The same has been suggested for lyrebirds, famous for their astonishing vocal mimicry (Zann & Dunstan, 2008).

Importantly, the mate choice hypothesis offers a plausible explanation for the origins of vocal learning (Nowicki & Searcy, 2014). Sexual signals are often thought to initially evolve by exploiting a pre-existing bias in the opposite sex (Ryan, 1998, 2018). A sensory bias for complex sounds can arise, for example, because, by being more varied, it prevents habituation and yields a higher response (Janik & Slater, 1997; Nowicki & Searcy, 2014). A classic test to verify whether a pre-existing sensory bias is responsible for the evolution of a sexual signal is to check whether females of a closely related species lacking the sexual signal share the preference for that signal (Ryan, 1998). In line with this, females of both wild zebra finches and wild white-backed munia prefer the more varied songs of conspecific males bred in captivity over the songs of their wild-type conspecifics (Okanoya, 2002; Tchernichovski et al., 1998). Exploitation of a pre-existing preference for complex song can thus explain initial selection for complex vocal learning in order to produce more complex song.

Sexual selection has been advocated as an explanation for the evolution of human language and music (Darwin, 1871; Fitch, 2010; Miller, 2000; Mithen, 2007). Darwin (1871, pp. 56-57) argued that "language owes its origin

to imitation", that is complex vocal learning, and that the evolution of complex vocal learning was initially driven by sexual selection. He asserted that "… some early progenitor of man, probably first used his voice in producing true musical cadences, that is in singing, as do some of the gibbon-apes at present day; and we may conclude from a widespread analogy, that this power would have been especially exerted during the courtship of the sexes, …" This hypothesis, namely that courtship singing was a precursor of human language, has been dubbed "musical protolanguage" (Fitch, 2006, 2010; Mithen, 2007). Empirical evidence for a role of mate choice in the evolution of music is limited but testing has only recently gained traction (Charlton, 2014; Charlton et al., 2012; Miller, 2000; 2001; Mosing et al., 2014: Ravignani, 2018).

A recent concern raised against the mate choice hypothesis of complex vocal learning is the finding that female song is more widespread than previously thought and even ancestral in birds (Garamszegi et al., 2007; Odom et al., 2014). Due to the traditional association between sexual selection and male-biased sexual dimorphism, the fact that in some groups both sexes sing has been deemed problematic for the mate choice origin hypothesis (Odom et al., 2014; Searcy & Nowicki, 2019). Similarly, the lack of sex differences regarding musical ability, production and perception has been taken as inconsistent with a role for mate choice in humans (Mehr et al., 2020; Savage et al., 2020). However, since groups where both sexes sing (humans, 90% of bird species) are predominantly socially monogamous (Lack, 1968; Stewart-Williams & Thomas, 2013), which is closely associated with biparental care (Cockburn, 2006), which in turn selects for both sexes to be choosy (Courtiol et al., 2016), mutual mate choice might have favored the evolution of vocal learning ability to attract mates in both sexes in these groups (Kokko & Johnstone, 2002). Indeed, paternal care and mutual mate choice are well-established in humans (Stewart-Williams & Thomas, 2013) and evidence indicates several uses of female birdsong including mate attraction (Riebel et al., 2019). Interestingly, the greater vasa parrot (*Coracopsis vasa*), for instance, demonstrates that female song can function to attract males that take up a provisioning role (Ekstrom et al., 2007). Hence, arguably, rather than that female song compromises the mate choice origin hypothesis, it expands its scope to include male mate choice for female song (e.g., Langmore et al., 1996), even though the latter is currently vastly understudied (Riebel et al., 2019).

In summary, based on a careful evaluation of immediate benefits, the most plausible explanations for the origins of complex vocal learning appear to be the mate choice and information-sharing hypotheses, despite the fact that several other functions likely operate, sometimes simultaneously, to currently maintain this trait (Nowicki & Searcy, 2014). Given the obvious advantages of complex vocal learning and the diversity of taxa that do have this ability, it is puzzling that it has not evolved in more species (Larsson & Abbott, 2018; Sewall et al., 2016). This

can only partly be explained by unevenly distributed research effort, as testified, for instance, by the fact that decades of intensive research into nonhuman primates yielded little evidence in favor of well-developed vocal learning. Therefore, it seems reasonable to assume that constraining factors should play a major role in explaining its infrequency (Jarvis, 2006).

Constraints

Regarding neurological constraints, it may be hypothesized that higher general cognition plays a role in the peculiar distribution of complex vocal learning among vertebrates. Yet, a telencephalic explanation does little to resolve the issue: nonhuman primates and many other highly encephalized and cognitively enhanced vertebrates, such as hyenas, lions, and dogs, lack complex vocal learning abilities (Larsson & Abbott, 2018; Sewall et al., 2016). Moreover, while complex vocal learning is a highly specialized mechanism which evolved by brain pathway duplication from an ancient motor learning pathway (Feenders et al., 2008; Jarvis, 2019), this motor learning pathway is a deeply homologous trait shared among many, if not all, vertebrate species (Feenders et al., 2008; Jarvis, 2019). Therefore, this cannot account for important disparities either, such as between extant human and great ape vocal learning abilities. On the contrary, as their material cultures show, great apes do have excellent motor learning skills in non-vocal domains (Fuhrmann et al., 2014).

The infrequency of complex vocal learning might be due to anatomical constraints. Specifically, it has been suggested that loss of air sacs and the presence of a permanently descended larynx in humans and additional intrinsic syrinx muscles in songbirds are required for flexible modifications of vocalizations (Fitch, 2018; Garcia et al., 2017). However, it has been shown that many vocal non-learning mammals descend their larynx as well when vocalizing. Moreover, other vocal non-learning mammals, such as lions and some ungulates, have independently evolved a permanently descended larynx. More likely, the descended larynx evolved to produce lower-formant frequencies, to acoustically exaggerate size (Fitch, 2018). With respect to birds, it has been found that syrinx muscle complexity does not correlate with vocal learning, but rather allows a vocal non-learner to produce a greater variety of innate sounds (Garcia et al., 2017). Finally, the larynx of a nonhuman monkey has been shown to be speech ready, indicating that the long-held view that vocal organ morphology is the culprit for the lack of complex vocal learning in nonhuman primates is incorrect (Fitch et al., 2016). Hence these previously proposed anatomical constraints on the vocal organ can unlikely account for the patchy taxonomical distribution of complex vocal learning (Fitch, 2000; 2017; Jarvis, 2019).

Ecological constraints that have been suggested might be more promising. Jarvis (2006) and Okanoya (2002) proposed that predation risk selects against vocal learning. Jarvis (2006) argues that complex male vocalizations

with more varied syntax may be attractive to predators for the same reason they are attractive to females, namely countering habituation, which makes them stand out from the background noise. This implies that complex vocal learners should have relatively few major predators. Top predators such as humans and killer whales, but also (adult) elephants, parrots and hummingbirds may fit this prediction (Jarvis, 2006). Okanoya (2002) showed that Bengalese finches that have been bred in captivity without human selection for singing behavior and without predators have more varied syntax than their White-backed Munia conspecifics from which they descended and which are still living in the wild. This cautiously suggests that predation pressure might indeed suppress elaboration of songs through complex vocal learning in the wild. However, phylogenetic comparative analyses should be carried out to test whether interspecies variation in complex vocal learning is indeed associated with predation risk (Münkemüller et al., 2012).

An especially promising ecological factor is the dimensionality of the environment. That is, environments with low dimensionality, such as the one-dimensional world of burrows and tunnels and two-dimensional dry land may impose a major constraint on complex vocal learning. Janik and Slater (1997) already noticed that a commonality shared among complex vocal learning groups is that they spend at least part of their lives swimming or flying in three-dimensional environments. A potential cause for this pattern, they note, is that it is more difficult for males to monopolize females in three-dimensional space and that this impacts their mating strategy and the probability to evolve complex vocal learning. In the next section I elaborate on this suggestion in the light of recent theoretical and empirical advances.

Three-dimensionality of the environment favors mate choice

Two distinct processes that constrain female mate choice are suggested to be significantly hampered in three-dimensional environments: male defense of territories occupied by females against rivalling males (i.e., male-male contest competition; dimensionality hypothesis, Puts, 2010) and male sexual coercion (coercion-avoidance hypothesis, Pradhan & van Schaik, 2009). As a result, it is expected that female choice exerts a comparatively strong selection pressure on males in three-dimensional environments (which, as discussed, may in some species subsequently select for male mate choice for female ornaments such as song). Hence, if the mate choice hypothesis of complex vocal learning were correct (Janik & Slater, 1997; Nottebohm, 1972; Nowicki & Searcy, 2014), an association between complex vocal learning and three-dimensionality of the environment across taxa would be expected.

The *dimensionality hypothesis* holds that the relative prevalence of sexual selection mechanisms - including mate choice - depends on the dimensionality of the physical mating environment (Puts, 2010). More precisely, in

the essentially one-dimensional environments of burrows and tunnels, and in two-dimensional environments such as dry land it is feasible for males to exclude same-sex competitors from mates or the resources necessary to attract them. As a result contest competition will be the main mechanism of sexual selection, typically leading to selection for secondary sexual characteristics that aid in intrasexual contests such as strength, large size and weapons (e.g., canines, horns and antlers) (Darwin, 1871; Rico-Guevara & Hurme, 2019). However, whereas the difficulty of defense increases linearly with the radius of the defense region in two dimensions, it increases with the square of this radius in three-dimensional environments. As a result, in three-dimensional environments (air, water, or trees), where there are too many in-routes for competitors, monopolization of females or the resources necessary to attract them becomes nearly impossible. Consequently, other sexual selection mechanisms, such as sperm competition and female mate choice, may be favored. Mate choice often leads to the evolution of ornamental traits to attract mates, ranging from morphological structures such as combs, elongated tails, and brightly colored beaks or patches of fur or skin to olfactory and acoustic courtship displays including songs (Andersson, 1994; Darwin, 1871). Puts (2010) suggests that the dimensionality of the mating environment explains substantial interspecific variation in contest competition, and consequently, other sexual selection mechanisms such as mate choice. Hence, the combination of the mate choice hypothesis of vocal learning and dimensionality hypothesis predicts a positive association between vocal learning and three-dimensionality, mediated by mate choice.

The *coercion-avoidance hypothesis* also predicts that in three-dimensional habitats female mate choice will exert stronger selection on males to become ornamental rather than weaponized (Pradhan & van Schaik, 2009), for a reason complementary to the dimensionality hypothesis. Pradhan and van Schaik (2009) argue that males are expected to use sexually dimorphic weaponry not only to displace other males, but also to overcome female preferences and thus acquire matings by force whenever they can. Females should therefore avoid coercive males and avoid using weaponry as a criterion for male quality wherever possible, and rely on male viability indicators that cannot be used to coerce females, namely ornaments. Pradhan and van Schaik (2009) suggest that the ability to freely move in three dimensions, such as air (flight) or water (swimming under water), or forest canopy (climbing and leaping), increases the female's escape ability relative to the two-dimensional terrestrial habitat. Therefore, they predict that, in addition to dissipation of contest competition, coercion avoidance will also favor the evolution of ornaments (such as songs) in three-dimensional environments.

Complex vocal learning in three-dimensional space

At the species level, an overwhelming majority (over 99% percent) of complex vocal learners currently inhabit three-dimensional environments. This is even the case when the speciose oscine songbirds are left out of the

equation, because it is also the case that the established complex vocal learning groups that do not live in 3D constitute just a few extant species (*Homo sapiens* and two or three elephant species). However, when comparing across species it is necessary to correct for phylogenetic relations (Harvey & Pagel, 1998). For instance, vocal learning probably evolved only once in the common ancestor of the thousands of oscine songbird species (n=1). Taking that into consideration, the habitats of complex vocal learning clades still bear out the predicted association to a significant extent. As the ensuing informal comparative analysis will show, six out of eight clades that have independently evolved complex vocal learning, inhabit three-dimensional environments and for several of them evidence exists that they produce courtship songs.

Flight is ancestral in birds (Voeten et al., 2018) and thus the three complex vocal learning bird groups (parrots, oscine songbirds and hummingbirds) have been competing for mates in three-dimensional, aerial environments. Correspondingly, mate choice has traditionally been regarded as the predominant mechanism of sexual selection among birds (Andersson, 1994; Emlen & Oring, 1977). Oscine songbirds (Beecher & Brenowitz, 2005) and hummingbirds (Araya-Salas et al., 2018; Araya-Salas & Wright, 2013) use - often complex - learned vocalizations to attract mates. However, with interesting exceptions such as the polyandrous greater vasa parrot, in which females sing complex songs to attract males (Ekstrom et al., 2007), and despite their ability to do impressive imitations of human song, many parrot species currently do not possess elaborate songs but instead use complex vocal learning to develop short, individually distinctive contact calls (Berg et al., 2012).

Bats are the only mammals capable of self-powered flight and thus the sole mammals that have been competing for mates in the three-dimensional environment of air just as birds (Altringham, 1996). A growing body of research suggests that vocal learning is a highly prevalent feature of bat behavior (Knörnschild, 2014; Prat et al., 2015; Vernes, 2017) and that songs are an important part of the courtship behavior of many bat species (Behr & Von Helversen, 2004; Puechmaille et al., 2014; Toth & Parsons, 2018).

Because they are obligatorily aquatic, cetaceans have necessarily been competing for mates in a three-dimensional environment as well (Uhen, 2007). Baleen whales use learned signals primarily in song (all sexed singers have been males). Song is usually heard during the breeding season and never in the summer season when the whales are feeding, which strongly suggests that baleen whale song is involved in mating (Whitehead & Rendell, 2012). Humpback whales, for instance, produce long bouts of complex song that can last for hours primarily during the breeding season. It has been suggested that sexual selection is the main driver of the evolution of these elaborate signals (Herman, 2017; Janik, 2014). Toothed whales, however, do not appear to produce song

to attract mates, but produce learned contact calls for recognition (e.g., signature whistles in bottlenose dolphins) (Sayigh et al., 2013).

Pinnipeds are semi-aquatic mammals, but, as Slater and Janik (1997) already noted, differences occur between phocids and otariids regarding the dimensionality of their mating environments and their mating strategies. Apart from the elephant seal, phocids breed at least partly on ice and copulate in the water. This three-dimensional environment makes it hard for males to coerce females into mating or to defend several females against rivals. Males are known to vocalize extensively in water during the breeding season, suggesting that they instead use their complex vocal learning capacities to attract females (Janik & Slater, 1997; Stirling and Thomas, 2003). This also is the case for walruses (Sjare and Stirling, 1996; Sjare et al., 2003). Otariids, such as sea lions, in contrast, breed on land and males often engage in fierce contests, defending harems against other males (Janik & Slater, 1997). Correspondingly, there is currently little evidence available for complex vocal learning in otariids, despite decades of captive study (Cook et al., 2013; Janik & Slater, 1997; Schusterman, 2008).

Two out of the eight confirmed complex vocal learning groups currently inhabit terrestrial environments: humans and elephants. Moreover, elephants are not known to produce courtship songs but rather have been argued to use their complex vocal learning capacities for contact calls that serve to maintain individual-specific bonds within changing social groupings as they live in complex fission-fusion societies (Poole et al., 2005). Human males and females, on the other hand, do produce ornamental vocalizations (namely in the form of vocal music), that might have been under mate choice selection (Charlton, 2014; Charlton et al., 2012; Miller, 2000; 2001; Mosing et al., 2014: Ravignani, 2018).

In summary, this informal comparative analysis generally supports the positive association between complex vocal learning and the dimensionality of the environment mediated by mate choice. Consistent with this hypothesis, oscine songbirds, hummingbirds, bats, seals and baleen whales are complex vocal learners that compete for mates in three-dimensional environments and have been reported to produce courtship songs. Parrots and toothed whales inhabit three-dimensional environments as well, but they are not known to sing courtship songs (with notable exceptions such as the greater vasa parrot), while the few elephant species and humans form an exception in that they are terrestrial complex vocal learners. These exceptions suggest that they may require a different explanation in terms of evolutionary function. However, in the next section I will assert that alternatives also face problems and that an explanation in terms of mate choice is possible for these exceptions as well, although tentatively.

Social cohesion vs. mate choice origin

Parrots and toothed whales inhabit three-dimensional environments, but they generally use their complex vocal learning abilities to produce contact calls rather than courtship songs. And since they live in complex fission-fusion societies, these taxa seem rather in line with Sewall et al.'s (2016) hypothesis that vocal learning originated for social cohesion. However, according to Nowicki & Searcy (2014), despite its obvious current role in these clades, it is unlikely that social cohesion explains its origins, because group specific signatures do not occur unless vocal learning has already evolved. Furthermore, Sewall et al.'s (2016) hypothesis entails that, after originating for social cohesion, complex vocal learning has been co-opted for courtship songs in clades that produce such vocalizations. However, if this were true, it would be expected that these singing clades also learn their contact calls. Yet, singing clades hummingbirds and baleen whales, for instance, have not been reported to modify the acoustic properties of their calls (Sewall, 2015). As Sewall (2015) notes this is consistent with the fact that these animals generally live more solitary than social lives. However, it contradicts Sewall et al.'s (2016) social cohesion origin hypothesis, because they are complex vocal learners, unless singing clades underwent a decrease in social complexity accompanied with the loss of vocal learning in the domain of contact calls. It might be possible, but it seems less likely than the reverse scenario, namely that complex vocal learning originated to produce varied courtship song and was later co-opted in some lineages for contact call learning where social complexity increased (Nowicki & Searcy, 2014).

Janik (2014) makes exactly this point with respect to whales. He suggests that, since singing is common in baleen whales where social structures are less complex, complex vocal learning in cetaceans may have evolved in a sexual selection context and was then later available for use in social negotiations when more complex social systems evolved in the toothed whales. Moreover, it is conceivable that ornamental song production largely disappeared in some taxa (e.g., parrots, toothed whales) as social complexity increased because mate attraction became redundant (e.g., due to increased proximity and frequency of interactions) and/or because songs became more contact call-like, as has been observed in budgerigars (Moravec et al., 2006). Whether and if learned contact call production evolved from learned courtship song production (or vice versa) could be confirmed with formal evolutionary transition analyses (cf. Mizuno & Soma, 2020).

A final possibility is that complex vocal learning originated for social cohesion functions in some taxa and for courtship in other taxa, although this would go against Nowicki and Searcy (2014) logical argument that social cohesion has a hard time explaining its origins. At any rate, formal phylogenetic comparative analyses could be used to resolve this issue by reconstruction of ancestral functions (Harvey & Pagel, 1998).

A role for ancestral three-dimensionality?

Elephants

Humans and elephants also do not appear to be in line with the 3D/mate choice hypothesis, as they are terrestrial mammals competing for mates on essentially two-dimensional dry land. Elephants also do not sing to attract mates. Rather - consistent with the dimensionality hypothesis - bulls fight over mates, especially during musth (Sukumar, 2003). Yet they do produce learned contact calls, presumably to mediate social relations since they live in complex fission fusion societies (Poole et al., 2005). It seems therefore reasonable to assume that elephant complex vocal learning also originated in elephants for this social cohesion function (Sewall et al., 2016). However, given that a social cohesion origin goes against Nowicki and Searcy's (2014) immediate benefit argument, it might be worthwhile to explore alternative explanations for complex vocal learning origin among elephants. As discussed, in contrast to social cohesion, mate attraction can offer an immediate benefit to the complex vocal learner even if no one else has the capacity yet. And since the evolution of ornaments to attract mates, including songs, is associated with three-dimensional environments (Pradhan & van Schaik, 2009; Puts, 2010), ancestral three-dimensionality would be consistent with a mate choice origin of elephant complex vocal learning. In fact, it is well established that elephants have gone through an aquatic – and thus three-dimensional – phase during their evolution. This might have offered elephants a window of opportunity to pick up complex vocal learning and later they may have maintained it through selection for social cohesion when they returned to dry land. Albeit admittedly very tentative at this stage, it might nonetheless be interesting to survey the evidence from diverse fields to explore this possibility.

Regarding phylogenetics, the closest extant relatives of elephants are the aquatic sea-cows (dugongs and manatees; order Sirenia) and the terrestrial hyraxes (order Hyracoidea) (Kellogg et al., 2007). Very few studies have been conducted on sea-cow and hyrax vocal learning behavior and abilities. One study found that manatee calls resemble those of their mothers, but could not determine whether this was due to vocal learning or genetic inheritance (Sousa-Lima et al., 2002). However, another study found that the calls produced by a calf, which was rejected by its mother and bottlefed in an isolated pen, resembled its mother's, whereas the vocal patterns of calls of twin calves, which were raised in a communal pen, demonstrated no similarities with the vocal patterns of their mother's. The authors consider the influence of vocal learning 'a likely explanation' as calves in captivity that are exposed to vocal templates of unrelated individuals might develop a vocal pattern different from the primarily inherited one (Sousa-Lima et al., 2008).

Male rock hyraxes sing complex songs that demonstrate geographical dialects with regard to syntax and syllable order. While it is not clear what sort of genetic mechanism could affect syllable order and syntax, Kershenbaum et al., (2012) deem it likely that individual and within-regional variation in song syntax and syllable order is due to the fact that dispersing males carry song features from their natal group, which are then repeated and learnt by hyraxes at the destination sites. Based on such observations, Kershenbaum et al. (2012) suggest that their songs might involve 'copying and generating novel vocalizations' and thus complex vocal learning. They regard imprecise copying or improvisation a likely scenario for maintaining similarity gradients as they have observed along such dispersal paths. And the lack of correlation between male vocal profile and their genetic relatedness within one site provides additional support for vocal learning. Hence, interestingly, elephants have among their closest living relatives both aquatic and potentially complex vocal learner animals.

Morphological, developmental, paleontological, and, more recently, molecular evidence for an aquatic – and thus three-dimensional - ancestry of elephants has been steadily accumulating and now constitutes a solid case (Gaeth et al., 1999; Liu et al., 2008; Mirceta et al., 2013). Reconstruction based on, among other things, net surface charge of myoglobin indicates that the last common ancestor of sea-cows, elephants, and hyraxes had diving capacities of the magnitude otherwise only observed in lineages of expert divers, such as cetaceans and pinnipeds (Mirceta et al., 2013). Hence, elephants might share with hyraxes their vocal abilities as well as an exceptional return from a secondarily aquatic to a terrestrial habitat, which would be consistent with a vocal learning origin in a common ancestor that lived in a three-dimensional environment.

Humans

With respect to humans, Nowicki and Searcy (2014) conclude, based on their careful evaluation of major hypotheses, that information sharing with kin and mate choice are the most plausible explanations for the origin of human vocal learning. While the information sharing hypothesis is considered plausible by several scholars, a potential weakness is that it would apply only to humans and not to other animals, as humans are the only animals in which learned vocal signals provide environmental information to others (Nowicki & Searcy, 2014). Reserving a special explanation uniquely for humans seems at odds with Darwin's (1871) continuity principle that asserts we should explain human evolution in terms of the same general processes we use to explain the behaviors of other biological entities, based on the fact that humans are products of the same evolutionary processes.

A weakness of the mate choice hypothesis, on the other hand, is that there are reasons to doubt that it has been impactful enough in *Homo sapiens* to drive the evolution of complex vocal learning. Traditionally, evolutionary

psychologists assume that mate choice has been the primary mechanism of sexual selection in humans, but new advancements rather suggest that contest competition has been the primary mechanism of sexual selection – at least in men. Phylogeny, the spatial and temporal clustering of mates and competitors, and anatomical considerations point this out. Men's traits are better designed for contest competition and coercion than for other sexual selection mechanisms; size, muscularity, strength, aggression, and the manufacture and use of weapons probably helped ancestral males win contests directly, and deep voices and facial hair signal dominance more effectively than they increase attractiveness (Puts, 2010; Saxton et al., 2016). This is consistent with the dimensionality and coercion avoidance hypotheses, since they predict that contest competition and coercion override mate choice, because mates are easier to monopolize in the essential two-dimensional mating environment of humans (Pradhan & van Schaik, 2009; Puts, 2010).

Yet, it is possible that human complex vocal learning originated in a time when humans inhabited an environment with higher dimensionality and monopolizations and coercion of females was less effective as a result of which (mutual) mate choice could freely select for ornaments such as complex song. When human ancestors subsequently shifted to a two-dimensional environment, they could have retained complex vocal learning abilities because it had already been co-opted for other uses such as information sharing and social cohesion. Alternatively, although it was somewhat weakened, mate choice might not have disappeared despite this habitat shift because of the infant dependency, paternal care and mutual mate choice that had already coevolved with it. Some comparative evidence is consistent with this scenario, which is, admittedly, quite speculative at this stage.

An option that might first come to mind in this context is the aquatic ape hypothesis (Morgan, 1997; Niemitz, 2010; Verhaegen, 2013). The basic premise of this hypothesis is that a semi-aquatic lifestyle has played a morphological role in the evolution of our human ancestors, leading to many of the major differences between humans and the great apes. However, to have the necessary impact regarding dimensionality of the mating environment, one would need to resort to an untenably strong version of the aquatic ape hypothesis. A weaker, 'waterside' version, is uncontroversial and backed up by plenty of empirical evidence (Finlayson, 2014), but would not provide the necessary three-dimensionality of the *mating* environment. Notice, for comparison, that even otariids such as sea lions, with clear morphological adaptations to an aquatic life, but that mate on land, do not have complex vocal learning, while closely related phocids, which compete for mates in the water, do (Janik & Slater, 1997; Schusterman, 2008; Stirling & Thomas, 2003).

An alternative possibility is ancestral arboreality. I first consider primates in general and then the human clade specifically. Puts' (2010) dimensionality hypothesis holds that three-dimensionality of the environment due to a

semi-arboreal lifestyle should correlate with reduced contest competition as in such a three-dimensional environment males cannot successfully defend their territory against rivals (Mitani & Rodman, 1979; Puts, 2010). The primary correlates of contest competition in primates are sexual dimorphism in canine and body size. Consistent with the dimensionality hypothesis, contests are more common among terrestrial primates, which exhibit greater body and canine size sexual dimorphism than arboreal and arboreal/terrestrial species (Clutton-Brock et al., 1977; Leutenegger & Cheverud, 1982; Plavcan & van Schaik, 1997; Puts, 2010). If contest competition and also coercion opportunities are reduced (Pradhan & van Schaik, 2009), other sexual selection mechanisms such as mate choice may kick in. In as far as mate choice selects for singing behavior, we should therefore expect a higher incidence of singing behavior among arboreal compared to terrestrial primates. And in effect, the only three nonhuman primate genera - Indri, Tarsius, Callicebus (titis) -, and one family – Hylobatidae (gibbons) - in which albeit unlearned - 'singing' species are found are all arboreal. Consistently with reduced contest competition and coercion, these are all relatively monomorphic primates (Geissmann, 2000). Moreover, evidence indicates that one of the functions of gibbon song may be mate attraction (Cowlishaw, 1996). Hence, it appears that 'song' among nonhuman primates is associated with arboreal three-dimensionality, as expected based on the dimensionality and coercion avoidance hypotheses. This raises the question whether it is conceivable that human song also has its origin in the trees.

Due to its apparent absence in nonhuman primates, the most parsimonious conclusion from phylogenetic comparative analysis is that complex vocal learning uniquely originated in the human lineage after the split from the lineage leading to the chimpanzees (Fitch, 2017; Petkov & Jarvis, 2012). The conventional view has been that song and speech only evolved several million years after this split, when the - terrestrial - genus *Homo* appeared. Yet, recent findings do not preclude the possibility that the vocal abilities underlying song and speech evolved before the appearance of *Homo* and even before human ancestors shifted from an arboreal to an exclusively two-dimensional terrestrial environment. It is now established that hominins maintained (semi-)arboreal lifestyles after the split from the chimpanzee lineage much longer than previously thought. Of particular significance in this respect are findings concerning *Ardipithecus* and in particular *Ardipithecus ramidus*. This species is situated near the base of the hominin clade and can as such inform us about human evolution at its beginnings (Lovejoy, 2009). Diverse lines of evidence indicate that *Ardipithecus* maintained a woodland-to-forest adaptation well into the Pliocene.

Furthermore, even the earliest Australopithecus species appear to have retained elements of woodland adaptation (White et al., 2015). There are some indications that *Australopithecus afarensis* was still to a large

extent tree-dwelling (Ruff et al., 2016). Thus, it might not have been a coincidence that Lucy, a fossil from this species with rock star fame, has recently been shown to have fallen from a tree (Kappelman et al., 2016). Based on a review of paleontological evidence, Finlayson (2014) concludes that at the earliest at the base of the Homo clade unconditional commitment to life on the ground appeared. Consequently, it seems that hominins, after the split with the lineage leading to chimpanzees, for an extended period of time, still competed for mates in a hemispherical, three-dimensional environment. It is therefore conceivable that this created a reasonably large window of opportunity for complex vocal learning to evolve driven by (mutual) mate choice for complex courtship songs.

Just as the extant 'singing' primates such as gibbons, the body and canine size of *Ardipithecus ramidus* was nearly monomorphic and male canine size reduced and 'feminized' in shape, consistent with reduced contest competition due to its (semi-)arboreal, tree-dimensional lifestyle (Suwa et al., 2009). Hence, in line with the dimensionality and coercion-avoidance hypotheses, a tree-dwelling human ancestor such as *Ardipithecus* appears a suitable candidate for the gibbon-like progenitor of man Darwin (1871) envisioned as the ancestor in which human complex vocal learning originated.

A problem for the arboreality explanation of complex vocal learning may be that, apart from the hypothesized human ancestors, there are no arboreal complex vocal learners. All other complex vocal learners' environments are three-dimensional due to being aerial or aquatic. It could be that effective three-dimensionality is generally lower in trees, because arboreal species are limited to moving along solid structures in their environment (Janik & Slater, 1997). It could also be that, since complex vocal learning is so rare and since much fewer vertebrate species live in trees than in the water and in the air, it is a mere statistical coincidence that there are no extant arboreal complex vocal learners. It could also be a combination of these two explanations.

If the infrequency of complex vocal learning is due to constraints, it would be worthwhile to explore how they may operate differently in humans and other primates. Sperm competition might be such a constraint. In effect, a trade-off between vocal production and sperm production has been found repeatedly, including in humans (Charlton & Reby, 2016; Dunn et al., 2015; Simmons et al., 2011). This trade-off is consistent with the fact that when females mate with more than one male, competition between males can continue after mating in the form of sperm competition. Allocation theory predicts (Parker, 1998), and several empirical findings confirm, that males should increase their investment in sperm production as sperm competition is increased, but it assumes that males face a trade-off between sperm production and other life-history traits such as attracting mates (e.g., Buzatto,

Roberts, & Simmons, 2015; Charlton & Reby, 2016; Dunn et al., 2015; Evans, 2010). Thus, investment in sperm may come at a cost of costly ornaments such as song production.

In line with this, the ratios of testes volume to body mass, a reliable indicator of sperm competition, are similar in humans and gibbons, consistent with the fact that both are 'singers', while the ratios are three times higher in chimpanzees (Pan) (Harcourt et al., 1981). Hence, despite environmental three-dimensionality and comparable levels of contest competition (Puts, 2010), human's closest extant relatives might not be able to imitate sounds due to having taken the evolutionary pathway of strong sperm competition.

In the same way as with elephants, hominin ancestors might have maintained their vocal learning capacity after the subsequent gradual shift to a terrestrial lifestyle in the genus Homo because it had already been co-opted for other functions (e.g., information sharing) besides mate attraction. Alternatively, mate choice might not have disappeared despite this habitat shift because of the infant dependency and paternal care that had already coevolved with it, which kept mutual mate choice going. Therefore, song may still play some role in human mate attraction until the present day, but musical capacity may be no longer as strongly selected relative to spoken language skills (Charlton, 2014; Mosing et al., 2014), which is currently deemed under powerful positive selection mainly for other purposes (Fitch, 2006).

In summary of this section, six out of eight complex vocal learning groups have been moving freely in the essentially three-dimensional environments of air because of flight (hummingbirds, oscine passerines, parrots and bats) and water because of an aquatic lifestyle (pinnipeds and cetaceans). Furthermore, in several of them this association is presently mediated by mate choice. Exceptions to this dimensionality rule are extant humans and elephants, which are currently terrestrial and thus essentially live in two-dimensional environments. However, it is conceivable, albeit tentative at this stage, that humans and elephants picked up complex vocal learning during a phase of ancestral three-dimensionality. Table 1 provides a summary. In the next section I consider vocal non-learning taxa inhabiting three-dimensional environments.

Table 1. Complex vocal learners and the dimensionality of their ancestral and current mating environment

Clades with Vocal Learning Species		*Ancestral Mating Environment*	*Dimensionality*	*Current Mating Environment*	*Dimensionality*
Mammals	Hominins	(Semi-)Arboreal		Terrestrial	
	Elephants	(Semi-)Aquatic		Terrestrial	2D
	Pinnipeds	(Semi-)Aquatic		(Semi-)Aquatic	
	Cetaceans	Aquatic		Aquatic	
	Bats	Aerial	3D	Aerial	
Birds	Oscine songbirds	Aerial		Aerial	3D
	Parrots	Aerial		Aerial	
	Hummingbirds	Aerial		Aerial	

Vocal non-learning taxa

Strong mate choice due to a three-dimensional mating environment is not a sufficient condition for the origin of complex vocal learning. Perhaps unsurprisingly, given its rarity, there are many more animal groups inhabiting three-dimensional environments (and, as expected, demonstrating strong mate choice) that do not appear to be complex vocal learners than groups that are. This may in part be due to the fact that only a fraction of species have been tested for this trait, but it is also likely due to various constraints, other than dimensionality (see section 'Constraints').

For several reasons fish are interesting to consider here. They are a paraphyletic group at the base of the vertebrate clade (all identified complex vocal learning species are vertebrates), they inhabit three-dimensional environments and with more than thirty thousand species, they are very speciose vertebrates (about three times as speciose as the already speciose bird clade). In line with dimensionality and coercion avoidance hypotheses, mate choice is relatively prevalent among fish as is the production of sounds to attract partners (Amorim et al., 2015). In contrast to terrestrial vertebrates or tetrapods, fish do not have a main vocal organ and instead have evolved multiple mechanisms to produce sounds (most commonly pulse series), many of which utilize sonic muscles that vibrate the swim bladder or the rubbing of bony elements (Fine & Parmentier, 2015; Ladich, 2014). However, many structures in and surrounding the vocal tract and lungs of tetrapods are homologous to the swim bladder and other respiratory anatomy in fish such as gill arches (Bass & Mckibben, 2003; Larsson & Abbott, 2018). Tetrapods and fish also share deep molecular commonalities. Genetic factors essential to vocalization and vocal learning in

tetrapod lineages, such as FoxP1 and FoxP2, are also present in fish where they regulate analogous social and learning behaviors (Condro & White, 2014; Scharff & Petri, 2011). Larsson and Abbott (2018) argue that vocal learning exapted auditory-motor circuits that evolved in fish for sonic awareness and motor entrainment. More specifically, they propose these circuits evolved via entrainment in the context of schooling behavior and respiratory-motor coupling, as incidental sounds of locomotion and respiration may have reinforced synchronization by communicating important spatial and temporal information between school-members. Hence, Larsson and Abbott (2018) suggest that brain pathways for vocal learning evolved in tetrapod lineages within strong constraints from fish. However, to date, no evidence exists of vocal learning in fish themselves. Yet, as Bass et al. (2015) point out, this could be due to the fact that it has remained vastly underexplored in fish.

Geographical variation in sound structure has been shown within several species (Fine, 1978; Phillips & Johnston, 2008; Tellechea et al., 2011), which is consistent with vocal learning, but does not constitute conclusive evidence. Evidence against vocal learning has been found in a cyprinid and in a cichlid species (Johnston & Buchanan, 2007; Longrie et al., 2008). Longrie et al. (2008) remark that very few fishes provide extensive posthatching parental care and, as such, the offspring have little opportunity to hear and learn sounds produced by the parents, which might impose an important constraint on the evolution of vocal learning. Thus future work might focus on the potential transmission of acoustic signals in groups that provide post-hatching parental care, such as sunfishes or cichlids (but see Johnston & Buchanan, 2007).

Several arboreal mammals (e.g., squirrels) are not known to learn their vocalizations either. Furthermore, many vocal non-learning amphibians and reptiles inhabit arboreal and aquatic three-dimensional ecosystems as well. But, while for instance frogs are well-known for their advertisement calls (e.g., Ryan & Rand, 1990), which is consistent with dimensionality and coercion-avoidance, little evidence for vocal learning has been reported in these taxa so far. A notable exception is the finding of vocal convergence in playback studies of white-lipped frogs, (*Leptodactylus albilabris*) (Lopez et al., 1988), indicating (at least) limited vocal learning ability.

While about half of bird species are complex vocal learners, and all of them are flying birds, this is largely due to the fact that oscine songbirds (suborder Passeri) constitute such a speciose clade. On a higher taxonomic level (family, order), complex vocal learners are in the minority among flying birds. Complex vocal learning has not been demonstrated in any birds of prey (e.g., owls, falcons, vultures), wading and water birds (e.g., gulls, waterfowl, shorebirds, grebes), cuckoos, and so on. It is thus clear that three-dimensionality is not sufficient for the evolution of complex vocal learning. Note however that neither is social complexity. Fish, ungulates, hyenas, nonhuman primates (including chimpanzees) and so on have often complex fission-fusion dynamics (Couzin &

Laidre, 2009; Kelley et al., 2011), yet have not been shown to imitate sounds. In conclusion, in many taxa living in three-dimensional environments there is little evidence of complex vocal learning, suggesting other factors may impose major constraints as well, but note as well that only a fraction of species have been tested.

Conclusions

Complex vocal learning is generally associated with three-dimensional environments, as expected based on coercion-avoidance, dimensionality and the mate choice hypotheses (Janik & Slater, 1997; Jarvis, 2006; Jarvis, 2019; Nottebohm, 1972; Nowicki & Searcy, 2014; Pradhan & van Schaik, 2009; Puts, 2010). Hummingbirds, oscine songbirds, bats, pinnipeds, and baleen whales are complex vocal learners currently inhabit three-dimensional environments that produce courtship songs. Parrots and toothed whales are also complex vocal learners inhabiting three-dimensional habitats, but are generally considered not to sing. Extant elephants and humans constitute the exceptional few species that have complex vocal learning abilities while inhabiting a basically two-dimensional, terrestrial habitat. Parrots, toothed whales and elephants produce learned contact calls instead of songs probably to promote social cohesion, which has led to the hypothesis that complex vocal learning originated for that function (Sewall et al., 2016). However, I have argued that it is more likely that complex vocal learning originally evolved for courtship and was later co-opted for social cohesion than the other way around. This is in line with Darwin's (1871) 'musical protolanguage' hypothesis that human complex vocal learning evolved for the production of ornamental courtship songs in a 'gibbon-like progenitor of man' and was later co-opted for low cost, language-like uses (Fitch, 2006, 2010; Mithen, 2007). Yet, multiple functional origins of complex vocal learning should not be excluded at this point, including information sharing with kin (only in humans).

Future research should verify with quantitative phylogenetic comparative methods, the effects of constraints such as three-dimensionality, predation risk and (newly proposed) sperm competition. Also whether mate choice, social cohesion, or other functions are responsible for the origin and/or maintenance of complex vocal learning, could be disentangled through, for example, ancestral state reconstruction and evolutionary transitions analyses.

Acknowledgements

I thank Sally Street, Krist Vaesen, Marcel Eens, and Gert Verpooten for discussion and comments on an initial version of the manuscript. I also thank the associate editor and two anonymous reviewers for their in-depth comments on previous versions of the manuscript.

Conflict of Interest

The author declares that he has no conflict of interest.

References

Abramson, J. Z., Hernández-Lloreda, M. 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: Biological Sciences*, *285*(1871), 20172171. https://doi.org/10.1098/rspb.2017.2171

Altringham, J. D. (1996). *Bats : biology and behaviour*. Oxford University Press.

Amorim, M. C. P., Vasconcelos, R. O., & Fonseca, P. J. (2015). Fish Sounds and Mate Choice. In F. Ladich (Ed.), *Sound communication in fishes*. Springer. https://doi.org/10.1007/978-3-7091-1846-7

Andersson, M. B. (1994). *Sexual selection*. Princeton University Press.

Araya-Salas, M., Gonzalez-Gomez, P., Wojczulanis-Jakubas, K., López, V., & Wright, T. F. (2018). Spatial memory is as important as weapon and body size for territorial ownership in a lekking hummingbird. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-20441-x

Araya-Salas, M., & Wright, T. (2013). Open-ended song learning in a hummingbird. *Biology Letters*, *9*(5), 20130625. https://doi.org/10.1098/rsbl.2013.0625

Arriaga, G., & Jarvis, E. D. (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language*, *124*(1), 96–116. https://doi.org/10.1016/j.bandl.2012.10.002

Bass, A. H., Chagnaud, B. P., & Feng, N. Y. (2015). Comparative Neurobiology of Sound Production in Fishes. In F. Ladich (Ed.), *Sound communication in fishes*. Springer. https://doi.org/10.1007/978-3-7091-1846-7

Bass, A. H., & Mckibben, J. R. (2003). Neural mechanisms and behaviors for acoustic communication in teleost fish. *Progress in Neurobiology*, *69*, 1–26. https://doi.org/10.1016/S0301-0082(03)00004-2

Beecher, M. D., & Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. *Trends in Ecology and Evolution*, *20*(3), 143–149. https://doi.org/10.1016/j.tree.2005.01.004

Behr, O., & Von Helversen, O. (2004). Bat serenades: Complex courtship songs of the sac-winged bat (Saccopteryx bilineata). *Behav Ecol Sociobiol*, *56*, 106–115. https://doi.org/10.1007/s00265-004-0768-7

Belyk, M., & Brown, S. (2017). The origins of the vocal brain in humans. *Neuroscience and Biobehavioral Reviews*, *77*, 177–193. https://doi.org/10.1016/j.neubiorev.2017.03.014

Berg, Delgado, S., Cortopassi, K. A., Beissinger, S. R., & Bradbury, J. W. (2012). Vertical transmission of learned signatures in a wild parrot. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1728), 585–591.

https://doi.org/10.1098/rspb.2011.0932

Blumstein, D. T., & Armitage, K. B. (1997). Does sociality drive the evolution of communicative complexity? A comparative test with ground-dwelling sciurid alarm calls. *American Naturalist*, *150*(2), 179–200. https://doi.org/10.1086/286062

Boncoraglio, G., & Saino, N. (2007). Habitat structure and the evolution of bird song: A meta-analysis of the evidence for the acoustic adaptation hypothesis. *Functional Ecology*, *21*(1), 134–142. https://doi.org/10.1111/j.1365-2435.2006.01207.x

Bradbury, J. W. ., & Vehrencamp, L. (2011). *Principles of animal communication* (second edition). Oxford University Press

Brockelman, W. Y., & Schilling, D. (1984). Inheritance of stereotyped gibbon calls. *Nature*, *312*(5995), 634–636. https://doi.org/10.1038/312634a0

Buzatto, B. A., Roberts, J. D., & Simmons, L. W. (2015). Sperm competition and the evolution of precopulatory weapons: Increasing male density promotes sperm competition and reduces selection on arm strength in a chorusing frog. *Evolution*, *69*(10), 2613–2624. https://doi.org/10.1111/evo.12766

Charlton, B. D. (2014). Menstrual cycle phase alters women's sexual preferences for composers of more complex music. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1784), 20140403–20140403. https://doi.org/10.1098/rspb.2014.0403

Charlton, B. D., Filippi, P., & Fitch, W. T. (2012). Do women prefer more complex music around ovulation? *PLoS ONE*, *7*(4), e35626. https://doi.org/10.1371/journal.pone.0035626

Charlton, B. D., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications*, *7*(1), 12739. https://doi.org/10.1038/ncomms12739

Clutton-Brock, T. H., Harvey, P. H., & Rudder, B. (1977). Sexual dimorphism, socionomic sex ratio and body weight in primates. *Nature*, *269*(5631), 797–800. https://doi.org/10.1038/269797a0

Cockburn, A. (2006). Prevalence of different modes of parental care in birds. *Proceedings. Biological Sciences*, *273*(1592), 1375–1383. https://doi.org/10.1098/rspb.2005.3458

Coleman, S. W., Patricelli, G. L., Coyle, B., Siani, J., & Borgia, G. (2007). Female preferences drive the evolution of mimetic accuracy in male sexual displays. *Biology Letters*, *3*(5), 463–466. https://doi.org/10.1098/rsbl.2007.0234

Condro, M. C., & White, S. A. (2014). Recent Advances in the Genetics of Vocal Learning. *Comparative Cognition & Behavior Reviews*, *9*, 75–98. https://doi.org/10.3819/ccbr.2014.90003

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), 412–427. https://doi.org/10.1037/a0032345

Courtiol, A., Etienne, L., Feron, R., Godelle, B., & Rousset, F. (2016). The Evolution of Mutual Mate Choice under Direct Benefits. *American Naturalist*, *188*(5). https://doi.org/10.1086/688658

Couzin, I. D., & Laidre, M. E. (2009). Fission-fusion populations. *Current Biology*, *19*(15), R633–R635. https://doi.org/10.1016/j.cub.2009.05.034

Cowlishaw, G. (1996). Sexual Selection and Information Content in Gibbon Song Bouts. *Ethology*, *102*(2), 272–284. https://doi.org/10.1111/j.1439-0310.1996.tb01125.x

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. John Murray.

Dunn, J. C., Halenar, L. B., Davies, T. G., Cristobal-Azkarate, J., Reby, D., Sykes, D., Dengg, S., Fitch, W. T., Knapp, L. A., Cristobal-Azkarate, J., Reby, D., & Sykes, D. (2015). Evolutionary Trade-Off between Vocal Tract and Testes Dimensions in Howler Monkeys. *Current Biology*, *25*(21), 2839–2844. https://doi.org/10.1016/j.cub.2015.09.029

Ekstrom, J. M. M., Burke, T., Randrianaina, L., & Birkhead, T. R. (2007). Unusual sex roles in a highly promiscuous parrot: The Greater Vasa Parrot *Caracopsis vasa*. *Ibis*, *149*(2), 313–320. https://doi.org/10.1111/j.1474-919X.2006.00632.x

Emlen, S. T., & Oring, L. W. (1977). Ecology, sexual selection, and the evolution of mating systems. *Science*, *197*(4300), 215–223. https://doi.org/10.1126/science.327542

Evans, J. P. (2010). Quantitative genetic evidence that males trade attractiveness for ejaculate quality in guppies. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1697), 3195–3201. https://doi.org/10.1098/rspb.2010.0826

Everett, C., Blasí, D. E., & Roberts, S. G. (2016). Language evolution and climate: the case of desiccation and tone. *Journal of Language Evolution*, *1*(1), 33–46. https://doi.org/10.1093/JOLE/LZV004

Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., Wada, K., Mouritsen, H., Jarvis, E. D., & Clayton, N. S. (2008). Molecular Mapping of Movement-Associated Areas in the Avian Brain: A Motor Theory for Vocal Learning Origin. *PLoS One*, *3*(3), e1768. https://doi.org/10.1371/journal.pone.0001768

Fine, M. L. (1978). Seasonal and geographical variation of the mating call of the oyster toadfish *Opsanus tau L.* *Oecologia*, *36*(1), 45–57. https://doi.org/10.1007/BF00344570

Fine, M. L., & Parmentier, E. (2015). Mechanisms of Fish Sound Production. In F. Ladich (Ed.), *Sound*

Communication in Fishes. (pp. 77–126). Springer. https://doi.org/10.1007/978-3-7091-1846-7_3

Finlayson, C. (2014). *The Improbable Primate: How Water Shaped Human Evolution*. Oxford University Press. https://global.oup.com/academic/product/the-improbable-primate-9780199658794?cc=be&lang=en&

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*, 258–267.

Fitch, W. T. (2004). Kin selection and "mother tongues": A neglected component in language evolution." In K. Oller & U. Griebel (Eds.), *Evolution of Communication Systems: A Comparative Approach* (pp. 275–296). MIT Press.

Fitch, W. T. (2005). The evolution of language: A comparative review. *Biology & Philosophy*, *20*(2–3), 193–203. https://doi.org/10.1007/s10539-005-5597-1

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

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

Fitch, W. T. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review*, *24*(1), 3–33. https://doi.org/10.3758/s13423-017-1236-5

Fitch, W. T. (2018). The Biology and Evolution of Speech: A Comparative Analysis. *Annual Review of Linguistics*, *4*(1), 255–279. https://doi.org/10.1146/annurev-linguistics-011817-045748

Fitch, W. T., De Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, *2*(12). https://doi.org/10.1126/sciadv.1600723

Freeberg, T. M. (2006). Social complexity can drive vocal complexity: Group size influences vocal information in Carolina chickadees. *Psychological Science*, *17*(7), 557–561. https://doi.org/10.1111/j.1467-9280.2006.01743.x

Fuhrmann, D., Ravignani, A., Marshall-Pescini, S., & Whiten, A. (2014). Synchrony and motor mimicking in chimpanzee observational learning. *Scientific Reports*, *4*. https://doi.org/10.1038/srep05283

Gaeth, A. P., Short, R. V, & Renfree, M. B. (1999). The developing renal, reproductive, and respiratory systems of the African elephant suggest an aquatic ancestry. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 5555–5558.

Garamszegi, L. Z., Pavlova, D. Z., Eens, M., & Møller, A. P. (2007). The evolution of song in female birds in Europe. *Behavioral Ecology*, *18*(1), 86–96. https://doi.org/10.1093/beheco/arl047

Garcia, S. M., Kopuchian, C., Mindlin, G. B., Fuxjager, M. J., Tubaro, P. L., & Goller, F. (2017). Evolution of Vocal Diversity through Morphological Adaptation without Vocal Learning or Complex Neural Control.

Current Biology, *27*(17), 2677-2683.e3. https://doi.org/10.1016/j.cub.2017.07.059

Geissmann, T. (2000). Gibbon songs and human music from an evolutionary perspective. In N. Wallin, B. Merker, & S. Brown (Eds.), *Origins of Music* (pp. 103–123). MIT Press.

Gould, S. J. (1991). Exaptation: A Crucial Tool for an Evolutionary Psychology. *Journal of Social Issues*, *47*(3), 43–65. https://doi.org/10.1111/j.1540-4560.1991.tb01822.x

Hansen, P. (1979). Vocal learning: its role in adapting sound structures to long-distance propagation, and a hypothesis on its evolution. *Animal Behaviour*, *27*, 1270–1271. https://doi.org/10.1016/0003-3472(79)90073-3

Harcourt, A. H., Harvey, P. H., Larson, S. G., & Short, R. V. (1981). Testes weight, body weight and breeding system in primates. *Nature*, *293*, 55–57.

Harvey, P. P., & Pagel, M. D. (1998). *The Comparative Method in Evolutionary Biology*. Oxford University Press.

Hayes, K. J., & Hayes, C. (1952). Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology*, *45*(5), 450–459. https://doi.org/10.1037/h0053609

Herman, L. M. (2017). The multiple functions of male song within the humpback whale (Megaptera novaeangliae) mating system: Review, evaluation, and synthesis. *Biological Reviews*, *92*, 1795–1818. https://doi.org/10.1111/brv.12309

Hindmarsh, A. M. (1984). Vocal mimicry in starlings. *Behaviour*, *90*, 302–324. https://doi.org/10.1163/156853984X00182

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, *3*(7), 272–279. https://doi.org/10.1016/S1364-6613(99)01333-9

Janik, V. M. (2014). Cetacean vocal learning and communication. *Current Opinion in Neurobiology*, *28*, 60–65. https://doi.org/10.1016/j.conb.2014.06.010

Janik, V. M., & Slater, P. J. B. (1997). Vocal Learning in Mammals. *Advances in the Study of Behavior*, *26*(C), 59–99. https://doi.org/10.1016/S0065-3454(08)60377-0

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*(1), 1–11. https://doi.org/10.1006/ANBE.2000.1410

Jarvis, E. (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science*, *5*(1), 5–14. https://doi.org/10.2326/osj.5.5

Jarvis, E. D. (2009). Bird song systems: Evolution. *Encyclopedia of Neuroscience*, 217–225. https://doi.org/10.1016/B978-008045046-9.00935-9

Jarvis, E., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y. W., Faircloth, B., Nabholz, B., & Al, E. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*(6215), 1126–1138. https://doi.org/10.1126/science.1251385

Jarvis, Erich D. (2019). Evolution of vocal learning and spoken language. *Science*, *366*(6461), 50–54. https://doi.org/10.1126/science.aax0287

Johnston, C. E., & Buchanan, H. M. (2007). Learned or innate production of acoustic signals in fishes: A test using a cyprinid. *Environmental Biology of Fishes*, *78*(2), 183–187. https://doi.org/10.1007/s10641-006-9087-3

Kappelman, J., Ketcham, R. A., Pearce, S., Todd, L., Akins, W., Colbert, M. W., Feseha, M., Maisano, J. A., & Witzel, A. (2016). Perimortem fractures in Lucy suggest mortality from fall out of tall tree. *Nature*, *537*(7621), 503–507. https://doi.org/10.1038/nature19332

Kelley, J. L., Morrell, L. J., Inskip, C., Krause, J., & Croft, D. P. (2011). Predation risk shapes social networks in fission-fusion populations. *PLoS ONE*, *6*(8), 24280. https://doi.org/10.1371/journal.pone.0024280

Kellogg, M. E., Burkett, S., Dennis, T. R., Stone, G., Gray, B. A., McGuire, P. M., Zori, R. T., & Stanyon, R. (2007). Chromosome painting in the manatee supports Afrotheria and Paenungulata. *BMC Evolutionary Biology*, *7*, 6. https://doi.org/10.1186/1471-2148-7-6

Kershenbaum, A., Ilany, A., Blaustein, L., & Geffen, E. (2012). Syntactic structure and geographical dialects in the songs of male rock hyraxes. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1740), 2974–2981. https://doi.org/10.1098/rspb.2012.0322

Killin, A. (2016). Reflections on imitation, vocal mimicry, and entrainment. *Aisthesis. Pratiche, Linguaggi e Saperi Dell'estetico*, *9*(2), 81–87. https://doi.org/10.13128/AISTHESIS-19418

King, S. L., & Janik, V. M. (2013). Bottlenose dolphins can use learned vocal labels to address each other. *Proceedings of the National Academy of Sciences*, *110*(32), 13216–13221. https://doi.org/10.1073/pnas.1304459110

Knörnschild, M. (2014). Vocal production learning in bats. *Current Opinion in Neurobiology*, *28*, 80–85. https://doi.org/10.1016/j.conb.2014.06.014

Knörnschild, M., Nagy, M., Metz, M., Mayer, F., & Von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biology Letters*, *6*(2), 156–159. https://doi.org/10.1098/rsbl.2009.0685

Kokko, H., & Johnstone, R. A. (2002). Why is mutual mate choice not the norm? Operational sex ratios, sex roles and the evolution of sexually dimorphic and monomorphic signalling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*(1419), 319–330. https://doi.org/10.1098/rstb.2001.0926

Kondo, N., & Watanabe, S. (2009). Contact calls: Information and social function. *Japanese Psychological Research*, *51*(3), 197–208. https://doi.org/10.1111/j.1468-5884.2009.00399.x

Lack, D. (1968). *Ecological adaptations for breeding in birds*. Methuen.

Ladich, F. (2014). Fish bioacoustics. *Current Opinion in Neurobiology*, *28*, 121–127. https://doi.org/10.1016/j.conb.2014.06.013

Laland, K. N., & Janik, V. M. (2006). The animal cultures debate. *Trends in Ecology and Evolution*, *21*(10), 542–547. https://doi.org/10.1016/j.tree.2006.06.005

Langmore, N. E., Davies, N. B., Hatchwell, B. J., & Hartley, I. R. (1996). Female song attracts males in the alpine accentor *Prunella collaris*. *Proceedings of the Royal Society B: Biological Sciences*, *263*(1367), 141–146. https://doi.org/10.1098/rspb.1996.0022

Larsson, M., & Abbott, B. W. (2018). Is the Capacity for Vocal Learning in Vertebrates Rooted in Fish Schooling Behavior? *Evolutionary Biology*, *45*(4), 359–373. https://doi.org/10.1007/s11692-018-9457-8

Lattenkamp, E. Z., & Vernes, S. C. (2018). Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Current Opinion in Behavioral Sciences*, *21*, 209–215. https://doi.org/10.1016/j.cobeha.2018.04.007

Leutenegger, W., & Cheverud, J. (1982). Correlates of sexual dimorphism in primates: Ecological and size variables. *International Journal of Primatology*, *3*(4), 387–402. https://doi.org/10.1007/BF02693740

Liu, A. G. S. C., Seiffert, E. R., & Simons, E. L. (2008). Stable isotope evidence for an amphibious phase in early proboscidean evolution. *Proceedings of the National Academy of Sciences*, *105*(15), 5786–5791. https://doi.org/10.1073/pnas.0800884105

Longrie, N., Fine, M. L., & Parmentier, E. (2008). Innate sound production in the cichlid *Oreochromis niloticus*. *Journal of Zoology*, *275*(4), 413–417. https://doi.org/10.1111/j.1469-7998.2008.00457.x

Lopez, P. T., Narins, P. M., Lewis, E. R., & Moore, S. W. (1988). Acoustically induced call modification in the white-lipped frog, Leptodactylus albilabris. *Animal Behaviour*, *36*(5), 1295–1308. https://doi.org/10.1016/S0003-3472(88)80198-2

Lovejoy, C. O. (2009). Reexamining human origins in light of *Ardipithecus ramidus*. *Science*, *326*(5949). https://doi.org/10.1126/science.1175834

Lupyan, G., & Dale, R. (2016). Why Are There Different Languages? The Role of Adaptation in Linguistic Diversity. *Trends in Cognitive Sciences*, *20*(9), 649–660. https://doi.org/10.1016/j.tics.2016.07.005

Maddieson, I., Bhattacharya, T., Smith, D. E., & Croft, W. (2011). Geographical distribution of phonological

complexity. *Linguistic Typology*, *15*(2), 267–279. https://doi.org/10.1515/LITY.2011.020

Marlowe, F. W. (2005). Hunter-gatherers and human evolution. *Evolutionary Anthropology*, *14*(2), 54–67. https://doi.org/10.1002/evan.20046

McComb, K., & Semple, S. (2005). Coevolution of vocal communication and sociality in primates. *Biology Letters*, *1*(4), 381–385. https://doi.org/10.1098/rsbl.2005.0366

Mehr, S., Krasnow, M., Bryant, G., & Hagen, E. (2020). Origins of music in credible signaling. *Behavioral and Brain Sciences*. https://doi.org/10.31234/osf.io/nrqb3

Meyer, J. (2004). Bioacoustics of human whistled languages: An alternative approach to the cognitive processes of language. *Anais Da Academia Brasileira de Ciencias*, *76*(2), 405–412. https://doi.org/10.1590/s0001-37652004000200033

Miller, G. F. (2000). Evolution of Human Music through Sexual Selection. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The Origins of Music* (pp. 329–360). MIT Press. https://doi.org/10.7551/mitpress/5190.003.0025

Miller, G. F. (2000). *The Mating Mind: How Sexual Choice Shaped the Evolution of the Human Nature*. Doubleday.

Mirceta, S., Signore, A. V., Burns, J. M., Cossins, A. R., Campbell, K. L., & Berenbrink, M. (2013). Evolution of mammalian diving capacity traced by myoglobin net surface charge. *Science*, *340*(6138). https://doi.org/10.1126/science.1234192

Mitani, J. C., & Rodman, P. S. (1979). Territoriality: The relation of ranging pattern and home range size to defendability, with an analysis of territoriality among primate species. *Behavioral Ecology and Sociobiology*, *5*(3), 241–251. https://doi.org/10.1007/BF00293673

Mithen, S. (2007). The Singing Neanderthals: The Origins of Music, Language, Mind and Body. Weidenfeld and Nicolson.

Mizuno, A., & Soma, M. (2020). Potential role of sensory bias in plumage pattern evolution: termite-eating and polka-dots in estrildid finches. *Ethology, Ecology and Evolution*, 1–13. https://doi.org/10.1080/03949370.2020.1803414

Moravec, M. L., Striedter, G. F., & Burley, N. T. (2006). Assortative Pairing Based on Contact Call Similarity in Budgerigars, Melopsittacus undulatus. *Ethology*, *112*(11), 1108–1116. https://doi.org/10.1111/j.1439-0310.2006.01267.x

Morgan, E. (1997). *The Aquatic Ape Hypothesis: The Most Credible Theory of Human Evolution*. Souvenir Press.

Mosing, M. A., Verweij, K. J. H., Madison, G., Pedersen, N. L., Zietsch, B. P., & Ullén, F. (2014). Did sexual

selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evolution and Human Behavior*, *36*(5), 359–366. https://doi.org/10.1016/j.evolhumbehav.2015.02.004

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, *3*(4), 743–756. https://doi.org/10.1111/j.2041-210X.2012.00196.x

Niemitz, C. (2010). The evolution of the upright posture and gait: A review and a new synthesis. *Naturwissenschaften*, *97*(3), 241–263. https://doi.org/10.1007/s00114-009-0637-3

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., & Jenner, K. C. S. (2000). Cultural revolution in whale songs: Humpbacks have picked up a catchy tune sung by immigrants from a distant ocean. *Nature*, *408*(6812), 537. https://doi.org/10.1038/35046199

Nottebohm, F. (1972). The Origins of Vocal Learning. *The American Naturalist*, *106*(947), 116–140. https://doi.org/10.1086/282756

Nowicki, S., & Searcy, W. A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology*, *28*, 48–53. https://doi.org/10.1016/j.conb.2014.06.007

Odom, K. J., Hall, M. L., Riebel, K., Omland, K. E., & Langmore, N. E. (2014). Female song is widespread and ancestral in songbirds. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms4379

Okanoya, K. (2002). Sexual Display as a Syntactic Vehicle: The Evolution of Syntax in Birdsong and Human Language through Sexual Selection. In A. Wray (Ed.), *The Transition to Language* (pp. 46–63). Oxford University Press.

Parker, G. A. (1998). Sperm Competition and the Evolution of Ejaculates: Towards a Theory Base. In T. Birkhead & A. P. Møller (Eds.), *Sperm Competition and Sexual Selection* (pp. 3–54). Academic Press. https://doi.org/10.1016/b978-012100543-6/50026-x

Payne, R. B. (1984). Sexual Selection, Lek and Arena Behavior, and Sexual Size Dimorphism in Birds. *Ornithological Monographs*, *33*, iii–52. https://doi.org/10.2307/40166729

Payne, R. B., Payne, L. L., and Woods, J. L. (1998). Song learning in brood-parasitic indigobirds Vidua chalybeata: Song mimicry of the host species. *Animal Behaviour* 55, 1537-1553.

Payne, R. B., Payne, L. L., Woods, J. L., and Sorenson, M. D. (2000). Imprinting and the origin of parasite-host species associations in brood-parasitic indigo-birds, *Vidua chalybeata*. *Animal Behaviour* 59, 69-81.

Peters, S., Derryberry, E. P., & Nowicki, S. (2012). Songbirds learn songs least degraded by environmental

transmission. *Biology Letters*, *8*(5), 736–739. https://doi.org/10.1098/rsbl.2012.0446

Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, *4*. https://doi.org/10.3389/fnevo.2012.00012

Phillips, C. T., & Johnston, C. E. (2008). Geographical divergence of acoustic signals in *Cyprinella galactura*, the whitetail shiner (Cyprinidae). *Animal Behaviour*, *75*(2), 617–626. https://doi.org/10.1016/j.anbehav.2007.06.022

Plavcan, J. M., & van Schaik, C. P. (1997). Intrasexual competition and body weight dimorphism in anthropoid primates. *American Journal of Physical Anthropology*, *103*(1), 37–68. https://doi.org/10.1002/(SICI)1096-8644(199705)103:1<37::AID-AJPA4>3.0.CO;2-A

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature*, *434*(7032), 455–456. https://doi.org/10.1038/434455a

Pradhan, G. R., & van Schaik, C. P. (2009). Why do females find ornaments attractive? The coercion-avoidance hypothesis. *Biological Journal of the Linnean Society*, *96*(2), 372–382. https://doi.org/10.1111/j.1095-8312.2008.01131.x

Prat, Y., Taub, M., & Yovel, Y. (2015). Vocal learning in a social mammal: Demonstrated by isolation and playback experiments in bats. *Science Advances*, *1*(2). https://doi.org/10.1126/sciadv.1500019

Prum. (2012). Aesthetic evolution by mate choice: Darwin's really dangerous idea. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1600), 2253–2265. https://doi.org/10.1098/rstb.2011.0285

Prum, R. O. (2010). The lande-kirkpatrick mechanism is the null model of evolution by intersexual selection: Implications for meaning, honesty, and design in intersexual signals. *Evolution*, *64*(11), 3085–3100. https://doi.org/10.1111/j.1558-5646.2010.01054.x

Prum, R. O. (2013). Coevolutionary aesthetics in human and biotic artworlds. *Biology & Philosophy*, *28*(5), 811–832. https://doi.org/10.1007/s10539-013-9389-8

Puechmaille, S. J., Borissov, I. M., Zsebok, S., Allegrini, B., Hizem, M., Kuenzel, S., Schuchmann, M., Teeling, E. C., & Siemers, B. M. (2014). Female Mate Choice Can Drive the Evolution of High Frequency Echolocation in Bats: A Case Study with *Rhinolophus mehelyi*. *PLoS ONE*, *9*(7), e103452. https://doi.org/10.1371/journal.pone.0103452

Puts, D. A. (2010). Beauty and the beast: Mechanisms of sexual selection in humans. *Evolution and Human Behavior*, *31*, 157–175. https://doi.org/10.1016/j.evolhumbehav.2010.02.005

Ravignani, A. (2018). Darwin, Sexual Selection, and the Origins of Music. *Trends in Ecology & Evolution*, *33*(10), 716–719. https://doi.org/10.1016/J.TREE.2018.07.006

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

Richman, B. (1978). The synchronization of voices by gelada monkeys. *Primates*, *19*(3), 569–581. https://doi.org/10.1007/BF02373317

Richman, B. (1987). Rhythm and melody in gelada vocal exchanges. *Primates*, *28*(2), 199–223. https://doi.org/10.1007/BF02382570

Rico-Guevara, A., & Hurme, K. J. (2019). Intrasexually selected weapons. *Biological Reviews*, *94*(1), 60–101. https://doi.org/10.1111/brv.12436

Ridgway, S., Carder, D., Jeffries, M., & Todd, M. (2012). Spontaneous human speech mimicry by a cetacean. *Current Biology*, *22*(20), R860--R861. https://doi.org/10.1016/j.cub.2012.08.044

Riebel, K., Odom, K. J., Langmore, N. E., & Hall, M. L. (2019). New insights from female bird song: Towards an integrated approach to studying male and female communication roles. *Biology Letters*, *15*(4). https://doi.org/10.1098/rsbl.2019.0059

Ríos-Chelén, A. A., Salaberria, C., Barbosa, I., Macías Garcia, C., & Gil, D. (2012). The learning advantage: Bird species that learn their song show a tighter adjustment of song to noisy environments than those that do not learn. *Journal of Evolutionary Biology*, *25*(11), 2171–2180. https://doi.org/10.1111/j.1420-9101.2012.02597.x

Roberts, G. (2010). An experimental study of social selection and frequency of interaction in linguistic diversity. *Interaction Studies*, *11*(1), 138–159. https://doi.org/10.1075/is.11.1.06rob

Ruff, C. B., Burgess, M. L., Ketcham, R. A., & Kappelman, J. (2016). Limb Bone Structural Proportions and Locomotor Behavior in A.L. 288-1 ("Lucy"). *PloS One*, *11*(11), e0166095. https://doi.org/10.1371/journal.pone.0166095

Ryan, M. J. (1998). Sexual selection, receiver biases, and the evolution of sex differences. *Science*, *281*(5385), 1999–2003. https://doi.org/10.1126/science.281.5385.1999

Ryan, M. J. (2018). *A taste for the beautiful: The evolution of attraction*. Princeton University Press.

Ryan, M. J., & Rand, A. S. (1990). The sensory basis of sexual selection for complex calls in the tungara frog, Physalaemus pustulosus (sexual selection for sensory exploitation). *Evolution*, *44*(2), 305–314. https://doi.org/10.1111/j.1558-5646.1990.tb05200.x

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*. https://doi.org/10.1017/S0140525X20000333

Saxton, T. K., Mackey, L. L., McCarty, K., & Neave, N. (2016). A lover or a fighter? Opposing sexual selection pressures on men's vocal pitch and facial hair. *Behavioral Ecology*, *27*(2), 512–519. https://doi.org/10.1093/beheco/arv178

Sayigh, L. S., Fellner, W., Janik, V. M., King, S. L., & Wells, R. S. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1757), 20130053–20130053. https://doi.org/10.1098/rspb.2013.0053

Scarl, J. C., & Bradbury, J. W. (2009). Rapid vocal convergence in an Australian cockatoo, the galah *Eolophus roseicapillus*. *Animal Behaviour*, *77*(5), 1019–1026. https://doi.org/10.1016/j.anbehav.2008.11.024

Scharff, C., & Petri, J. (2011). Evo-devo, deep homology and FoxP2: Implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1574), 2124–2140. https://doi.org/10.1098/rstb.2011.0001

Schmidt, K. L., Moore, S. D., MacDougall-Shackleton, E. A., & MacDougall-Shackleton, S. A. (2013). Early-life stress affects song complexity, song learning and volume of the brain nucleus RA in adult male song sparrows. *Animal Behaviour*, *86*(1), 25–35. https://doi.org/10.1016/j.anbehav.2013.03.036

Schusterman, R. J. (2008). Vocal Learning in Mammals with Special Emphasis on Pinnipeds. In D. K. Oller & U. Griebel (Eds.), *Evolution of Communicative Flexibility* (pp. 41–70). MIT Press. https://doi.org/10.7551/mitpress/9780262151214.003.0003

Searcy, W. A. (2019). Animal communication, cognition, and the evolution of language. *Animal Behaviour*, *151*, 203–205. https://doi.org/10.1016/j.anbehav.2019.03.001

Searcy, W. A., & Nowicki, S. (2019). Birdsong learning, avian cognition and the evolution of language. *Animal Behaviour*, *151*, 217–227. https://doi.org/10.1016/j.anbehav.2019.01.015

Sewall, K. B. (2015). Social complexity as a driver of communication and cognition. *Integrative and Comparative Biology*, *55*(3), 384–395. https://doi.org/10.1093/icb/icv064

Sewall, K. B., Young, A. M., & Wright, T. F. (2016). Social calls provide novel insights into the evolution of vocal learning. *Animal Behaviour*, *120*, 163–172. https://doi.org/10.1016/j.anbehav.2016.07.031

Simmons, L. W., Peters, M., & Rhodes, G. (2011). Low Pitched Voices Are Perceived as Masculine and Attractive but Do They Predict Semen Quality in Men? *PLoS ONE*, *6*(12), 29271.

https://doi.org/10.1371/journal.pone.0029271

Sjare, B., & Stirling, I. (1996). The Breeding-Behavior of Atlantic Walruses, *Odobenus rosmarus rosmarus*, in the Canadian High Arctic. *Canadian Journal of Zoology* 74, 897-911.

Sjare, B., Stirling, I., and Spencer, C. (2003). Structural variation in the songs of Atlantic walruses breeding in the Canadian High Arctic. *Aquatic Mammals* 29, 297-318.

Slabbekoorn, H., & den Boer-Visser, A. (2006). Cities Change the Songs of Birds. *Current Biology*, *16*(23), 2326–2331. https://doi.org/10.1016/j.cub.2006.10.008

Soma, M., & Garamszegi, L. Z. (2011). Rethinking birdsong evolution: Meta-analysis of the relationship between song complexity and reproductive success. *Behavioral Ecology*, *22*(2), 363–371. https://doi.org/10.1093/beheco/arq219

Sousa-Lima, R. S., Paglia, A. P., & da Fonseca, G. A. B. (2008). Gender, age, and identity in the isolation calls of Antillean manatees (*Trichechus manatus manatus*). *Aquatic Mammals*, *34*(1), 109–122. https://doi.org/10.1578/AM.34.1.2008.109

Sousa-Lima, R. S., Paglia, A. P., & Da Fonseca, G. A. B. (2002). Signature information and individual recognition in the isolation calls of Amazonian manatees, *Trichechus inunguis* (Mammalia: Sirenia). *Animal Behaviour*, *63*(2), 301–310. https://doi.org/10.1006/ANBE.2001.1873

Stafford, K. M., Lydersen, C., Wiig, Ø., & Kovacs, K. M. (2018). Extreme diversity in the songs of Spitsbergen's bowhead whales. *Biology Letters*, *14*, 20180056. https://doi.org/10.1098/RSBL.2018.0056

Stansbury, A. L., & Janik, V. M. (2019). Formant Modification through Vocal Production Learning in Gray Seals. *Current Biology*, *29*(13), 2244-2249.e4. https://doi.org/10.1016/j.cub.2019.05.071

Stewart-Williams, S., & Thomas, A. G. (2013). The Ape That Thought It Was a Peacock: Does Evolutionary Psychology Exaggerate Human Sex Differences? *Psychological Inquiry*, *24*, 137–168. https://doi.org/10.1080/1047840X.2013.804899

Stirling, I., & Thomas, J. A. (2003). Relationships between underwater vocalizations and mating systems in phocid seals. *Aquatic Mammals*, *29*(2), 227–246. https://doi.org/10.1578/016754203101024176

Stoeger, A. S., & Manger, P. (2014). Vocal learning in elephants: Neural bases and adaptive context. *Current Opinion in Neurobiology*, *28*, 101–107. https://doi.org/10.1016/j.conb.2014.07.001

Stoeger, A. S., Mietchen, D., Oh, S., De Silva, S., Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology*, *22*(22), 2144–2148. https://doi.org/10.1016/j.cub.2012.09.022

Sukumar, R. (2003). *The Living Elephants: Evolutionary Ecology, Behaviour, and Conservation*. Oxford University Press.

Suwa, G., Kono, R. T., Simpson, S. W., Asfaw, B., Owen Lovejoy, C., & White, T. D. (2009). Paleobiological implications of the *ardipithecus ramidus* dentition. *Science*, *326*(5949), 94–99. https://doi.org/10.1126/science.1175824

Symington, M. M. F. (1990). Fission-fusion social organization in Ateles and Pan. *International Journal of Primatology*, *11*(1), 47–61. https://doi.org/10.1007/BF02193695

Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P., & Ghazanfar, A. A. (2015). The developmental dynamics of marmoset monkey vocal production. *Science*, *349*(6249), 734–738. https://doi.org/10.1126/science.aab1058

Takahashi, D. Y., Liao, D. A., & Ghazanfar, A. A. (2017). Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys. *Current Biology*, *27*(12), 1844-1852.e6. https://doi.org/10.1016/j.cub.2017.05.004

Tchernichovski, O., Schwabl, H., & Nottebohm, F. (1998). Context determines the sex appeal of male zebra finch song. *Animal Behaviour*, *55*(4), 1003–1010. https://doi.org/10.1006/anbe.1997.0673

Tellechea, J. S., Norbis, W., Olsson, D., & Fine, M. L. (2011). Calls of the black drum (*Pogonias cromis:* Sciaenidae): geographical differences in sound production between northern and southern hemisphere populations. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, *315A*(1), 48–55. https://doi.org/10.1002/jez.651

Toth, C. A., & Parsons, S. (2018). The high-output singing displays of a lekking bat encode information on body size and individual identity. *Behavioral Ecology and Sociobiology*, *72*(7). https://doi.org/10.1007/s00265-018-2496-4

Tyack, P. L. (2020). A taxonomy for vocal learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20180406. https://doi.org/10.1098/rstb.2018.0406

Uhen, M. D. (2007). Evolution of marine mammals: Back to the sea after 300 million years. *The Anatomical Record*, *290*(6), 514–522. https://doi.org/10.1002/ar.20545

Verhaegen, M. (2013). The Aquatic Ape Evolves: Common Misconceptions and Unproven Assumptions About the So-Called Aquatic Ape Hypothesis. *Human Evolution*, *28*(3–4), 237–266.

Vernes, S. C. (2017). What bats have to say about speech and language. *Psychonomic Bulletin and Review*, *24*(1), 111–117. https://doi.org/10.3758/s13423-016-1060-3

Verpooten, J., & Eens, M. (in press). Singing is not associated with social complexity across species. *Behavioral*

and Brain Sciences.

Voeten, D. F. A. E., Cubo, J., de Margerie, E., Röper, M., Beyrand, V., Bureš, S., Tafforeau, P., & Sanchez, S. (2018). Wing bone geometry reveals active flight in Archaeopteryx. *Nature Communications*, *9*(1), 923. https://doi.org/10.1038/s41467-018-03296-8

Weinstein, D., Launay, J., Pearce, E., Dunbar, R. I. M., & Stewart, L. (2016). Singing and social bonding: changes in connectivity and pain threshold as a function of group size. *Evolution and Human Behavior*, *37*, 152–158. https://doi.org/10.1016/j.evolhumbehav.2015.10.002

White, T. D., Lovejoy, C. O., Asfaw, B., Carlson, J. P., & Suwa, G. (2015). Neither chimpanzee nor human, Ardipithecus reveals the surprising ancestry of both. *Proceedings of the National Academy of Sciences*, *112*(16), 4877–4884. https://doi.org/10.1073/pnas.1403659111

Whitehead, H., & Rendell, L. (2012). *The Cultural Lives of Whales and Dolphins*. University of Chicago Press.

Wilkinson, G. S. (2003). Social and vocal complexity in bats. In F. de Waal & P. L. Tyack (Eds.), *Animal Social Complexity: Intelligence, Culture and Individualized Societies* (pp. 86–112). Harvard University Press.

Winter, P., Handley, P., Ploog, D., & Schott, D. (1973). Ontogeny of Squirrel Monkey Calls Under Normal Conditions and Under Acoustic Isolation. *Behaviour*, *47*(3–4), 230–239. https://doi.org/10.1163/156853973X00085

Zann, R., & Dunstan, E. (2008). Mimetic song in superb lyrebirds: Species mimicked and mimetic accuracy in different populations and age classes. *Animal Behaviour*, *76*(3), 1043–1054. https://doi.org/10.1016/j.anbehav.2008.05.021