

ESSAY

# Vocal learning: Beyond the continuum

Pedro Tiago Martins<sup>1,2\*</sup>, Cedric Boeckx<sup>2,3</sup>

**1** Section of General Linguistics, Universitat de Barcelona, Barcelona, Spain, **2** University of Barcelona Institute of Complex Systems (UBICS), Barcelona, Spain, **3** Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain

\* [pt.martins@ub.edu](mailto:pt.martins@ub.edu)



## Abstract

Vocal learning is the ability to modify vocal output on the basis of experience. Traditionally, species have been classified as either displaying or lacking this ability. A recent proposal, the vocal learning continuum, recognizes the need to have a more nuanced view of this phenotype and abandon the yes–no dichotomy. However, it also limits vocal learning to production of novel calls through imitation, moreover subserved by a forebrain-to-phonatory-muscles circuit. We discuss its limitations regarding the characterization of vocal learning across species and argue for a more permissive view.

## OPEN ACCESS

**Citation:** Martins PT, Boeckx C (2020) Vocal learning: Beyond the continuum. *PLoS Biol* 18(3): e3000672. <https://doi.org/10.1371/journal.pbio.3000672>

**Published:** March 30, 2020

**Copyright:** © 2020 Martins, Boeckx. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** PTM acknowledges support from the Portuguese Foundation for Science and Technology (FCT - [www.fct.pt](http://www.fct.pt)) in the form of a PhD fellowship (grant SFRH/BD/131640/2017). CB acknowledges the financial support from the Spanish Ministry of Economy and Competitiveness ([www.mineco.gob.es](http://www.mineco.gob.es); grant FFI2016-78034-C2-1-P), a Marie Cure International Reintegration Grant from the European Union (<https://cordis.europa.eu>; grant PIRG-GA-2009-256413), research funds from the Fundació Bosch I Gimpera ([www.fbg.ub.edu](http://www.fbg.ub.edu)), the MEXT/JSPS Grant-in-Aid for Scientific Research on Innovative Areas 4903 (<https://www.jsps.go.jp/english/e-grants/>) (Evolvingistics: JP17H06379), and from the Generalitat de Catalunya (<https://web.gencat.cat/>; 2017-SGR-341). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Introduction

Humans (and no other species) have language. An important component for language is speech, which refers to the auditory/vocal medium we use to communicate linguistic units among ourselves, and apart from auditory perception and processing, it requires a vocal tract with a wide range of possibilities, such as ours (but not necessarily exactly like ours [1]), and the capacity for vocal learning. Vocal learning broadly construed is the ability to modify vocal output on the basis of experience. Unlike language, however, vocal learning is not unique to humans: it is present in several, distantly related species. There are different ways in which species modify their vocalizations (see [Box 1](#)). This could involve either a modification of an aspect of vocalization (as long as it can be shown that such a modification is learned from experience, usually auditory but not necessarily so [2]) or the production of novel vocalizations altogether.

Vocal learning is indeed a very productive area of study across disciplines [6]. Species that display vocal learning abilities are a relevant source of information on the nature and evolution of language in humans, chiefly regarding phonological aspects [7]. However, not everyone agrees on what constitutes vocal learning as a phenotype, and this greatly affects how work on vocal learning is carried out.

## The “canonical” list of vocal learners

There is a general trend in the literature (e.g., [8]) that limits vocal learning only to species that can produce novel calls through imitation, subserved by a direct connection between the

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** Am, nucleus ambiguus; DM, dorsal medial nucleus of the midbrain; KJ, Kuypers/Jürgens; NCBI, National Center for Biotechnology Information; PAG, periaqueductal gray; VLC, vocal learning continuum hypothesis; VTA, ventral tegmental area.

## Box 1. Simple vocal learning typology

When looking at vocal abilities, there are ways of teasing apart which behaviors require some form of learning and which don't. There is a major split between vocalizations that are innate, whose employment does not depend on experience, and those that do require experience and go beyond the innate repertoire. The former kind of vocal behavior is common to most animals. It includes crying and laughing [3], for example, and does not require experience (though see [4] for how experience influences prosodic aspects of crying). The latter kind is less widespread, and it consists broadly of three subtypes, following [5]:

- ability to associate a sound with a behavioral response (example: dog [*Canis familiaris*] response to human commands)
- ability to learn the context in which a vocalization can be used (example: vervet monkey [*Chlorocebus pygerythrus*] vocalizations in response to predators)
- ability to modify vocalizations on the basis of experience (example: birdsong), which can converge or diverge from a model

The behavior that interests us here is vocal production learning, which is what most researchers refer to when they refer to vocal learning. But what constitutes vocal production learning as a phenotype is far from agreed upon by researchers, both in contrast to the other subtypes and on its own. This naturally affects which species are considered capable of it.

forebrain and phonatory muscles (e.g., the larynx in mammals or syrinx in birds). This has led to a canonical list of vocal learning species. It comprises three bird orders and some mammals. The birds—by far the longest list when counting individual species (in the thousands [9])—comprise songbirds (Passeriformes), parrots (Psittaciformes), and hummingbirds (Trochiliformes) [10, 11]. The mammals include humans, some cetaceans [12, 13], pinnipeds [14, 15], elephants [16], and bats [17, 18]. Birds are considered closest to humans in vocal learning abilities, even though they are phylogenetically the most remote. Humans are the only primate uncontroversially considered to be vocal learners, whereas nonhuman primates are considered of little relevance in this regard. Refinements to this list usually consist of looking inward for finer distinctions in the families already established (e.g., bats [18] or parrots [19]), and rarely outward.

It is possible that this focus on imitation and novel vocalizations is due to it being the clearest case of something being “transmitted” and then “learned.” That is, it could be that for those who put a premium on imitation (e.g., [8]), learned (as opposed to “innate”) entails that there has to be imitation of something that wasn't there before in any form (in the repertoire), as opposed to improvisation or other ways in which sounds in a repertoire can change (for example, through social feedback or modifications of aspects of calls that do not entail an entirely novel output).

It is also possible that the appeal of “neuro-reductionism” (to virtually equate a behavioral phenotype with a neural implementation [20]), might have had an influence in the establishment of this take on vocal learning. The existence of a direct connection from forebrain to

phonatory muscles, allowing for fine control of those structures, is indeed an appealing idea on which to build.

There is, however, empirical evidence of vocal learning abilities in other species outside of the canonical list. Such evidence is usually behavioral and not an attempt to show direct fore-brain control of phonatory muscles. Indeed, there is work questioning that such a connection is a necessary condition for vocal learning in the first place [21, 22].

## The vocal learning continuum and beyond

In light of this, we think it worth discussing the “vocal learning continuum hypothesis” (VLC) [11], which categorizes species along a continuum of increasing vocal learning complexity. This is a valuable idea that goes against the traditional dichotomous view of vocal learning, according to which a species is either definitely a vocal learner or not at all. However, it too relies on production of novel calls through imitation, subserved by forebrain control of phonatory muscles, to determine the distribution of vocal learning abilities across species. Such an approach is therefore not representative of the diversity of vocal learning behavior across the animal kingdom [23, 24]. This diversity pertains not only to species for which there is recent evidence of vocal learning but also to the “well-established” vocal learning species, namely birds [25].

Indeed, we find that the way species learn to produce their communication signals should form the set of criteria that makes a species a vocal learner. The exact nature of the vocalizations and the neurobiology are of course extremely important, and they will allow for much more precise evolutionary work, but one must not lose sight of the fact that vocal learning is a behavioral phenotype, with learning as the most striking aspect.

Given this state of affairs, two ways offer themselves for future studies on vocal learning: (1) perpetuating the bifurcation between canonical and “negligible” vocal learning species or (2) turning attention to the behaviors observed and assessing them in the context of a broad sense of vocal learning, as opposed to dismissing them on neurobiological grounds alone.

Recently proposed frameworks and reflections also show some concern with this question and call for more wide-ranging perspectives on vocal learning (e.g., [26–28]).

In the remainder of the present paper, we start by going over the VLC and point out some limitations. We then point to evidence from “noncanonical” species that reinforces these limitations and conclude by outlining an extension to the VLC highlighting approaches to vocal learning that can help overcome them.

## Limitations of the vocal learning continuum

The VLC proposes that species can be placed along a continuum, yielding a gradual as opposed to dichotomous classification [11, 29–31]. The categories in the VLC are as follows: vocal non-learners, limited vocal learners, moderate vocal learners, complex vocal learners, and high vocal learners.

The motivation for the VLC is that some species cannot be clearly categorized as nonvocal learners or vocal learners (in the all-or-nothing sense), with the mouse (*Mus musculus*) being such a case: they seem to have some form of song (ultrasonic vocalizations), but it is not clear whether they are learned or innate. Some aspects of it, however, seem to be contingent on social feedback, which highlights the role of experience. In addition, species with very impressive but not identical vocal learning abilities, such as songbirds versus parrots, can also more safely be placed a notch apart so that they are not equated and the nuances that distinguish them are not lost.

Although the VLC is a very welcome and important proposal for the study of vocal learning, we find that it has some limitations, which must be overcome in order to achieve a full(er) picture of the vocal learning morphospace and an understanding of its evolutionary history.

Some of the limitations of the VLC are of a conceptual nature, and some are empirical. The conceptual limitations are independent of what the VLC is actually about and, instead, have to do with the validity of establishing a two-dimensional model of a complex trait, which had to evolve. The empirical limitations have to do with applying the model to vocal learning specifically and how its predictions don't pan out, for different reasons.

## Bidimensionality

Because vocal learning categories are determined by the existence and strength of a particular brain circuit in the VLC, this makes it a bidimensional system (see [28] for some recent discussion on the same issue).

The particular brain circuit is, to put it in simple terms, a direct connection from the forebrain to phonatory muscles, and it is thought to be present in some form in canonical vocal learning species. The Kuypers/Jürgens (KJ) hypothesis [8, 32] posits that such a connection is necessary for the kind of motor control that is required for vocal learning, and the VLC tacitly relies on it. This idea had already been made popular before (e.g., [33], among others), but it was perhaps made more widespread in the work of Jarvis (e.g., [11, 34]), and Fitch (e.g., [8, 35]), who named the hypothesis after two scientists who made important contributions to primate neurobiology [36–38]. In other words, even though the VLC is a more nuanced conception of how to ascribe vocal learning across species because it allows intermediate steps, it is still limited in the sense that it has the forebrain-to-phonatory-muscles connection as the sole predictor and allows for variation only in that dimension. It is relevant here to recognize the role of a direct forebrain-to-phonatory-muscles connection as a necessary ingredient in the VLC; proponents of the VLC are of course well aware that it alone cannot explain away vocal learning as a whole. Other abilities and traits are involved, such as auditory learning [11], but the VLC is not concerned with them.

An analogy to the bidimensional nature of the VLC would be a slider in a physical machine or a computer program that controls a parameter, and by sliding it back and forth, the output is changed. In this case, the parameter would be the strength of a forebrain-to-phonatory-muscles connection, and the output would be “less” or “more” vocal learning. If the slider is at position 0, we get no vocal learning. If the slider is at the maximum value, we get “high-end” vocal learning.

There are two ways in which this bidimensionality is problematic. The first is that it leaves out capacities and constraints at other levels of analysis [18, 24, 28], which might or might not go hand in hand with this brain circuit. This is well captured by the following questions, taken from [28]: (1) What makes a species a vocal learner? (2) When is vocal learning employed? (3) How can vocal learning be expressed by the organism? (4) Who (else) is capable of vocal learning? And (5) why did vocal learning evolve?

Recent empirical evidence shows that species that do not or are thought to not have relevant forebrain control of phonatory muscles can be vocal learners. This brings to the fore other ways in which species may achieve what is in effect vocal learning behavior. Testing of species whose vocal learning capacities are unknown or supposed not to exist still yields surprises.

The other problem is that if a goal (or even the main goal) of comparative work is to derive information about the evolution of traits and species, we cannot expect a single aspect (e.g., a single genetic change or a single brain connection) to offer a realistic picture of how the trait evolved [39]. Even if the empirical evidence established that only vocal learning species in any

one sense consistently have a certain brain connection and vice versa, evolutionary aspirations would still require a more complex explanation. In the realm of complex traits, there is always a cascade of effects with far-reaching implications [40]. It is also the case that even homologous behaviors don't necessarily share a neural mechanism: there can be genetic changes affecting circuitry with no change in behavior [41].

### More on brain wiring

The discovery of particular wiring (see [Box 2](#)) made it possible to attempt a principled, brain-based separation of strictly innate calls in a way that's shared among all mammals tested from calls that are controlled volitionally. However, even in the very strict sense of learning of novel vocalizations through imitation, it is not known beyond doubt that this is a necessary condition. For example, there are reports of learned, voiced calls in the orangutan (*Pongo* spp., a

#### Box 2. Two major pathways

There are two major pathways believed to be specifically involved in vocal behavior: a general, "primal" one that is associated with all vocalizing animals and, in addition, a more specific one that is associated with vocal learners.

The primal pathway goes from the anterior cingulate cortex to the PAG, to the reticular formation of the pons and medulla, and from there to the phonatory neurons [3]. It seems that the PAG pathway is not involved in vocal motor coordination but, instead, is responsible for initiation and intensity of what is in effect a vocal reaction. It is not involved in its patterning.

Besides this pathway, used for "reactive" or "affective" vocalizations, it is hypothesized that vocal learners also have a direct connection from the laryngeal motor cortex to the nucleus ambiguus (Am) and, from there, to the phonatory muscles. In birds, similar pathways are thought to exist. There is a connection from the dorsal medial nucleus of the midbrain (DM) to the 12th nerve nucleus, which controls the syrinx. This is the vocalization pathway analogous to the PAG pathway in, say, humans. In vocal learning birds, there is also a connection from the robust nucleus of the arcopallium to the 12th nerve nucleus [46]. Nonvocal learning birds are thought to not have such connections (e.g., pigeons [*Columbia livia*] [47], but evidence is scarce). This direct telencephalic connection in birds is analogous to the cortical connection in humans.

Not much is known about the presence of these connections in some of the families included in the canonical list of vocal learners (e.g., cetaceans and pinnipeds).

This association between medial pathways and innate vocalizations, on the one hand, and cortical pathways and vocal learning, on the other, has become established in the literature, but the claim made by the KJ hypothesis is not without challenges [21, 22]. Most relevant here are perhaps the criticisms by Lameira [22] because they are presented in light of comparative evidence. One argument has to do with attribution: the work by Kuypers and Jürgens does not show or entail what the hypothesis states. For example, Kuypers [36] is assumed to have shown that great apes did not have the required fore-brain-to-larynx connection, when in reality, he did in fact identify it in a chimpanzee (*Pan troglodytes*) subject, and Jürgens [38] used monkeys and not great apes in his work. This casts some doubt on our understanding of direct vocal control in chimpanzees and, potentially, other primates. The second argument has to do with evidence against what

the hypothesis predicts: nonhuman primates should not in any way display vocal learning. Yet evidence for primate vocal learning is accumulating (see, e.g., [22, 24, 48]). We go into more detail in section S1\_Text. Evidence in the opposite direction also exists: mice seem to have the required machinery, yet they are not vocal learners in the KJ sense [29]. From a neurobiological point of view, this should mean that either this particular connection is not necessary in principle for vocal learning or that nonhuman primates actually have it and that interpretations of the few data on this matter are incorrect. The third argument has to do with the very mechanical requirements the KJ hypothesis put forward for vocal learning, which rely heavily on vocal fold control. Also in the formalization of the VLC, this is assumed explicitly: “Vocal learning is the ability to modify the spectral and syntactic composition of vocalizations generated by the vocal organ (larynx in mammals or syrinx in bird)” [29]. This requirement leaves out supralaryngeal vocal production—equivalent to voiceless consonants in humans. These vocalizations, which in humans are the most widespread [49], involve the control of several structures above the “vocal organ,” such as the lips and jaw, and are used as well to expand the vocal repertoire. This might seem like a minor point, but it is worth emphasizing that human language, through speech, makes use of both voiced and voiceless sounds in all known languages. It is also the case that whispered speech, for the most part supralaryngeal, is intelligible, and there is evidence for the use of different acoustic cues in the absence of fundamental frequency [50]. Direct control of phonatory muscles—which produce voiced sounds—alone will leave a great deal unexplained. It has been suggested that, because control of these supralaryngeal structures is clearly present in nonhuman primates, laryngeal control is the extra neurobiological ingredient (a “derived trait” or autapomorphy) that made humans vocal learners [32, 51], but as far as we can assess, this only says something about the sound source and ultimately the acoustics, not about learning and, therefore, not about how ancient or widespread the ability would be in nonhuman species.

species that supposedly lacks the relevant connection [42]). It is also not clear whether the connection is sufficient (within reason) either; mice (*M. musculus*) apparently have the circuit but do not produce novel sounds through imitation [29, 30], and perhaps more interestingly, recent work shows that female zebra finches (*Taeniopygia guttata*), which do not produce learned song, have “male-like” song pathways [43], so the narrative is not totally compelling. Furthermore, there is work showing the involvement of other structures and pathways in the learning of vocal behavior in a relevant manner, such as the cerebellum [44], the periaqueductal gray (PAG) [2], or the ventral tegmental area (VTA) [45]. It is also not entirely clear why vocal learning, a phenotype whose most interesting aspect is arguably the learning part, must be limited to a certain kind of vocalization, namely the kind that requires fine control of the phonatory muscles (what is usually referred to as “phonation”). Moreover, it is important to bear in mind that any one connection does not exist in isolation; each brain region involved will be part of several other connections, each with its own complex evolutionary history.

It thus seems that, although phonatory muscle control is obviously a very useful ability, relying on the KJ hypothesis alone might not give us a good indication of the basis of vocal learning and how widespread the phenotype is.



## Complexity considerations

The VLC also aims to represent various degrees of vocal learning complexity. But as we will see on at least three counts, it does not do so satisfactorily.

Bengalese finches (*Lonchura striata domestica*) are the domesticated strain of the white-rumped munia (*L. striata*). Domestication has been claimed to increase vocal learning complexity: if a “wild” species is already a good vocal learner, it becomes a more complex vocal learner after undergoing domestication [52]. In the case of the Bengalese finch, for example, this happens despite the fact that this bird species was not bred for its song. It is possible that imitation—crucial per the KJ hypothesis and, concomitantly, the VLC—could actually be detrimental to syntactic complexity. Compared with their wild counterparts, Bengalese finches display higher unpredictability and syntactic complexity in their song because not only do they imitate their tutors (partially) but they also improvise, resulting in what is, in effect, lower imitation fidelity [53].

The full “classic” circuit of vocal learning involves a posterior pathway for vocalization and an anterior pathway for learning [34]. Besides those pathways, parrots were discovered to have a “shell” song system in addition to the “core” song system in all vocal learning birds [19]. A larger shell system relative to the core system is associated with parrot species that have more “complex” vocal learning abilities, and vice versa. Moreover, this shell system has mostly intercortical connections, as opposed to the direct connection to the motor neurons characteristic of the core system. It seems, then, that parrot species with a larger shell system have an edge in the VLC, but this edge is not related to the direct connection the VLC rests on. This poses a challenge to the VLC as it stands because it requires the addition of an extra factor (say, adjacent “song/speech” nuclei with intercortical connections or even just “strong intercortical connections”), opening way to a much less restrictive VLC because more factors would be added as needed for capturing differences between species, departing from the bidimensionality we have already alluded to. Indeed, Chakraborty and Jarvis [54] acknowledge it might not be straightforward to reconcile the core/shell system with the VLC.

Finally, in the VLC, humans alone are considered high-end vocal learners, whereas parrots are classified at a level just below, referred to as complex vocal learners. This is purportedly because of the higher syntactic complexity in human vocalizations, but this does not rest on the criteria for categorizing species along the VLC (presence and strength of direct connection to the phonatory organ and imitation). Language complexity need not even be instantiated in vocal behavior; it is well established that the linguistic capacity is the same in sign language (see [55]). It could be that, indeed, humans are the most advanced vocal learner, but this is not possible to discern from the criteria used in the VLC. It might have more to do with the process of cultural transmission and not with anything “vocal” [56]. In a manner similar to Bengalese finches, it has been hypothesized that the increased prosociality that characterizes domestication allows for the jump in complexity to take place (see [56] for discussion).

## A more permissive view

Recent work has a more wide-ranging view of what constitutes a vocal learning species and of what plays a role in it. This, we contend, is necessary in order to extend the idea of the VLC and overcome its limitations.

## Imitation and de novo vocalizations are not the whole story

An important step, in our view, is to adopt a view of vocal learning behavior not necessarily focused on imitation.

The production of de novo vocalizations (new in a species repertoire) and, furthermore, doing so through imitation is often taken as the golden standard when assessing vocal learning abilities. This is problematic because imitation is one possible means of displaying vocal learning behavior. Indeed, diverging from imitation is also a common phenomenon in vocal and cultural development [23].

Perhaps a more productive conception of vocal learning is looking at learned vocal behavior as having to be acquired in some manner over developmental time, especially in contingent ways (that is, dependent on experience and not a “certainty” given the initial state of the organism).

There are interesting cases that illustrate vocal development by diverging from the tutor song; that is, by the countering of or lack of imitation. Infant marmosets, for example, develop vocal learning abilities through social reinforcement from parents, not imitation. This leads to more control of the vocal apparatus, which allows them to produce lower entropy calls [24, 57].

Canaries (*Serinus canaria domestica*) trained on atypical song imitate it at first but, when reaching maturity, shape it into the species-specific song they were never exposed to [58]. Another well-known example is the de novo emergence of zebra finch song not by imitation but, instead, by the approximation of wild-type song over a couple of generations by birds reared in isolation, with no exposure to singing tutors [59].

Evidence of this kind is good indication that vocal learning is not driven (solely) by imitation and that vocal learning ability is characterized also by behaviors that suppress imitation.

### Evidence from species outside of the canonical vocal learners list

Opening up to more-permissive definitions of vocal learning goes hand in hand with opening up to the study of more species. A decent amount of evidence for vocal learning outside of the canonical list has been put forward, especially in recent years, with primates as the most representative of this trend, and some work on rodents. They moreover deserve special attention because there is resistance to taking this kind of evidence into account. Other species are more quickly accepted, perhaps because they employ imitation, and neurobiological information on these species is given a great deal of importance, given its scarcity. A good example of this is the African elephant (*Loxodonta africana*), which quickly entered the accepted list of vocal learning species [16].

The logistic difficulties in keeping and studying larger species as opposed to birds and other (usually smaller) species might also bias positions against, say, primate vocal abilities, leading to a situation in which absence of evidence might be mistaken for evidence of absence. Rodents present challenges on their own (e.g., several species produce ultrasonic vocalizations, which pose further challenges, and there is a lot of interindividual variation [60]). Studies on birds have unsurprisingly dominated vocal learning research (see data in [28]). Great ape language acquisition projects (e.g., [61]) might also have contributed to this state of affairs, given their varying goals and approaches, which usually had to do with finding some form of human language, as well as difficulty in interpreting their results. In [Box 3](#), we summarize some evidence that we think deserves, at the very least, attention if vocal learning is to be understood as a phenotype that's characterized by the learned modification of calls, with the exact nature of the calls being an important but not (dis)qualifying feature. A more complete (yet not exhaustive) list can be checked in [S1\\_Text](#).

We believe that evidence of the kind we review here has only been neglected because of its nonconformity with the KJ hypothesis. We find that one way of getting a full picture of vocal learning is placing the focus on observing the behavior, without preconceptions of what should

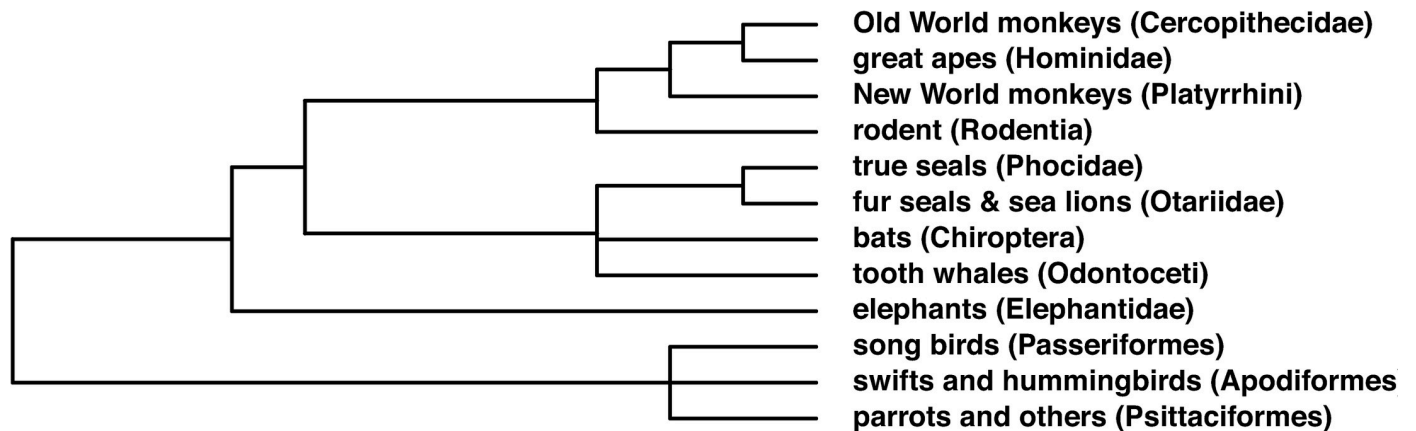


### Box 3. Evidence of vocal learning abilities in species outside of the canonical vocal learners list

Primates and rodents are not usually considered to be vocal learners, yet they display behavioral traits that fall within vocal learning in some sense. This is at odds with what circuitry is thought to be required for vocal learning behavior under the KJ hypothesis and the vocal learning continuum and warrants a rethinking of what is really known about the neurobiology of vocal learning. Moreover, some of these species are as well understood as others that do count as vocal learners in literature, warranting in this case a rethinking of the motivations for including some species but not others in the canonical list. Turning first to primates, both monkeys and apes display relevant behavior. Marmosets, a New World monkey who engages in turn taking [62], uses different proportions of affiliative call types depending on social distance [63], as well as loudness relative to physical distance [64]. They can also convey identity through aspects of their calls [65]. Their calls change from infancy into adulthood, much like in humans, and limiting parent feedback disrupts this development [66]. Several Old World monkeys display relevant vocal learning abilities. Diana monkeys show call converge in social interactions [67]. Campbell's monkeys, also a turn-taking species [68], sequence the sounds in their repertoire in a nonrandom way in different situations [69]. Rhesus monkeys have a juvenile period of volitional vocalizing, disappearing once adulthood is reached [70]. Apes show striking vocal learning abilities. Orangutans can learn voiced calls [42, 71] and whistles [72, 73]. They also employ “instrumental gesture calls,” whereby they volitionally use their hands or leaves in front of their mouth to lower the maximum frequency of calls [74, 75]. Gorillas [76] and chimpanzees [77] have also been shown to display vocal learning behavior. Turning now to rodents, there is promising evidence for vocal learning as well. Mice, who produce complex ultrasonic vocalizations, display variation in syllable type, which can distinguish between individuals [78]. They have also been shown to require feedback to maintain certain features of their song [30] and of changes in song development [79]. There is also Alston's mice, who engage in vocal bouts, which because of their length and patterning, have been deemed worthy of being called song [80, 81]. See [S1 Text](#) for an expanded list of species and abilities.

allow it, and then proceeding with the mechanisms. As put by Krakauer and colleagues [20]: “The neural basis of behavior cannot be properly characterized without first allowing for independent detailed study of the behavior itself.”

The relationship between a neural structure and a behavior is not one of explanation of the behavior [20]. This is the case even if that relationship is consistent, which in the case of vocal learning and according to evidence we reviewed, it might not be. Although some neuronal implementation will of course be in place, there is no indication that the vocal learning phenotype can be equated with a particular one given that there is not even a consensus on what the behavior encompasses. In the face of paradoxical evidence (e.g., primates displaying vocal behavior they are not “supposed to” have), rejecting the behavioral evidence instead of revising the neural hypothesis will not lead to understanding of the trait. It is in this sense that we think it is important to have behavior as the entry point to the study of vocal learning. If the VLC is



**Fig 1. Tree of vocal learners with a focus on behavior.** Differences between this tree and the canonical tree are reflected in the presence of primates and rodents. The inclusion of whole families or orders is made under the assumption that all members thereof are at least worth studying and is not a claim about their actual vocal learning abilities, which are an empirical question. Common and scientific names from NCBI (<https://www.ncbi.nlm.nih.gov/taxonomy>). Tree built with taxize R package [82]. NCBI, National Center for Biotechnology Information.

<https://doi.org/10.1371/journal.pbio.3000672.g001>

extended beyond the specific neural substrate that is taken to allow the direct control of the vocal organ, we could gain a better understanding of the phenotype.

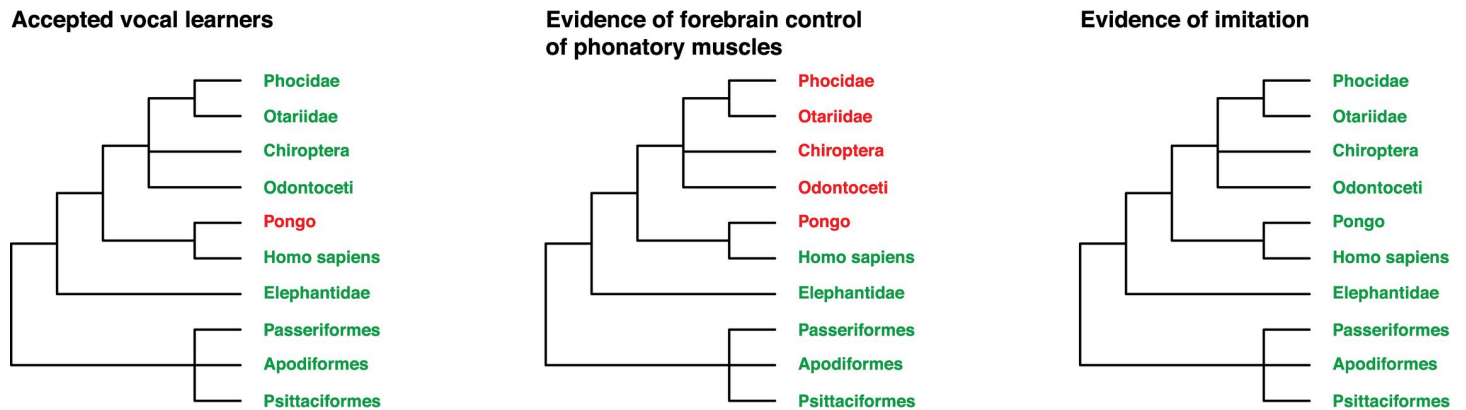
### Tree of vocal learners with a focus on behavior

Taking all the evidence available into account, and placing the focus in the behavior observed, we believe a more accurate “vocal learners list” (albeit with some tentative cases) would be the one we sketch in Fig 1.

We can see in this new list that it is possible to reduce the gap between us and the other vocal learners in a principled way. Although a direct forebrain–larynx connection is maybe not shared, there is much that is shared: similar patterns of early postnatal vocal development [66], volition [24], both voiceless and even voiced calls [42], socially reinforced vocal production, etc.

The canonical list of vocal learners, although much more manageable, is in effect a list of species for which there is, on the one hand, evidence of imitation and, on the other hand, evidence of direct connection from forebrain to phonatory muscles or an assumption of its existence (Fig 2, left). Assumption of its existence relies on two other assumptions: that this circuit is crucial for vocal learning and that nonhuman primates cannot have this circuit. But the fact of the matter is that there is no demonstration of this circuit for some species routinely considered vocal learners (Fig 2, center) that do show vocal imitation (Fig 2, right). In terms of evidence, there is nothing separating, say, orangutans from seals: there is evidence of imitation for both species, yet only one is an accepted vocal learner. Although one could argue about the strength of the evidence for some species over others, as well as ease of elicitation or perceived quality of the behavior, this disparity in the way different species are categorized seems to be stipulated by the KJ hypothesis and therefore warrants further reflection.

In previous work, we suggested the term “sound production” learning as opposed to vocal (production) learning as what might be a better term for defining the capacity we are interested in here, given the association in the literature of vocal with the phonatory muscles [83]. This might allow for a more encompassing definition, regardless of the mechanics involved. This would dilute a distinction that, as a characterization of the behavior, is not very relevant: if a species can change its repertoire, be it through imitation or not, the exact structures of the



**Fig 2. Comparison of simplified phylogenies of vocal learning: The canonical list with the addition of the orangutan (*Pongo*) for contrast.** Key: green indicates species or orders considered to conform to the criterion on each tree, and red indicates species or orders that are considered not to. Left: list of accepted vocal learners. Center: list of species for which there is evidence of a direct connection between forebrain and phonatory muscles. Right: list of species for which there is (some) evidence of vocal imitation.

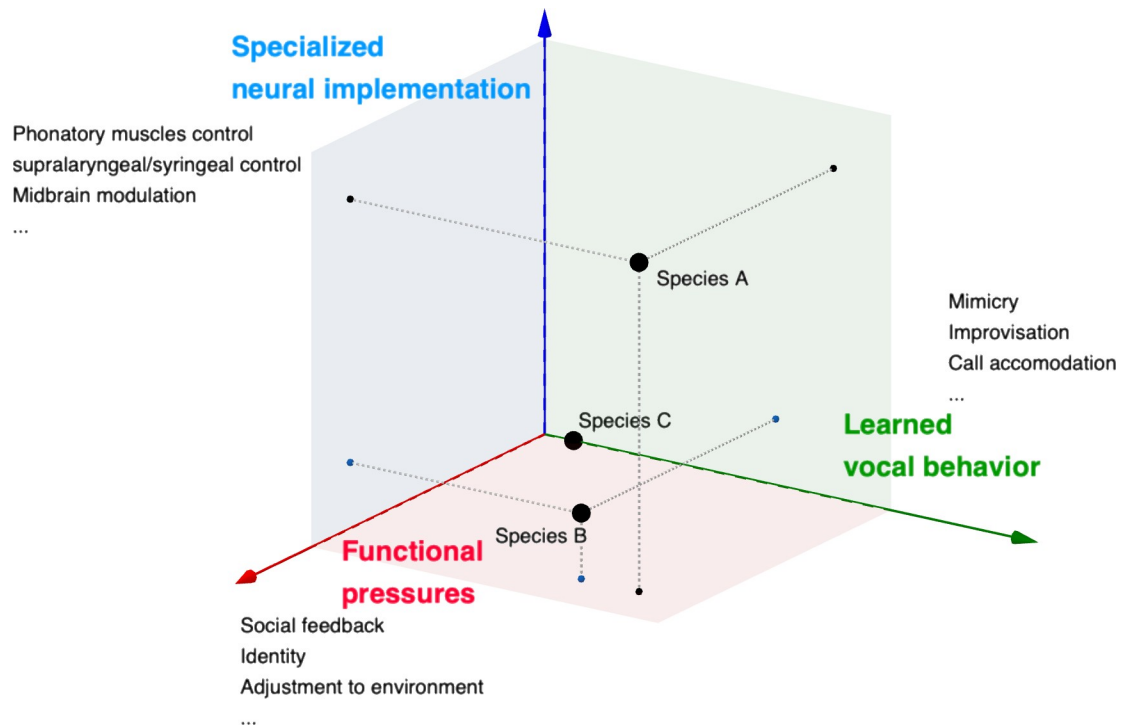
<https://doi.org/10.1371/journal.pbio.3000672.g002>

vocal tract that are used to doing so are not grounds for a big divide in the classification of the behavior, at least not with regard to learning. This is similar to what happens with the exact brain structures used in different vocal learning species: the fact that birds have no cortex and no larynx but, instead, telencephalon and syrinx does not warrant a strong divide as far as behavior is concerned, and indeed, birds are considered the prime model species for studying vocal learning in humans. Using the term sound production learning could lead to the inclusion of sound sources not limited to the vocal tract or other orofacial structures. In our view, it depends on the degree to which vocal learning is grounded in behavior and learning as opposed to the sound sources and the pathways expected to control them.

### Vocal learning contiguum: An outline

Focusing on just one measure of any one cognitive trait (that is, a “two-dimensional” continuum [28]), seems to be an attempt to find neat cognitive phylogenies (in the sense of Fitch and colleagues [35]) for what is a complex behavior (see [84] for discussion). This becomes an easier task if vocal learning is reduced to a single circuit because it allows one to conceive of single events (nodes in a phylogeny) that confer the behavior to a species and its descendants (e.g., a whole order of birds). If other factors are considered, however, different cognitive phylogenies could be devised. If we ask all the questions posed by Lattenkamp and Vernes [28] for each species, we will see that there will be gaps, but we will see as well that each species provides answers to at least some of the questions. This is therefore in our view an adequate list given all the evidence.

In Fig 3, we outline an extension of the VLC, which we call the vocal learning contiguum, to capture the notion of a space of neighboring and overlapping factors, as opposed to a linear scale (as in the VLC). In this conception, vocal learning is understood as a morphospace, and a species can be represented as displaying vocal abilities of a certain type without a necessary association with either a specific neural implementation or a specific set of functional pressures. If groups of species congregate in particular areas of the morphospace, one could take this as a good indication of which factors help shape them as vocal learners and to which degree. This is in a way a simplification of the several factors that contribute to a complex behavior. A more realistic picture would comprise several dimensions. Here, for representational purposes, we collapse them into three: evidence of vocal learning behavior in a broad sense, evidence of a specialized neuronal implementation, and evidence of functional pressures



**Fig 3. Outline of a vocal learning contiguuum.** Three main dimensions are considered: learned vocal behavior (green), functional pressures (red), and specialized neural implementation (blue). Examples are given to different factors that can be included in each of these dimensions. Black dots represent hypothetical species placed in the morphospace according to evidence. Species A displays vocal learning abilities and is thought to have a particular neuronal implementation in place, as well as subject to considerable functional pressures that help shape the behavior. Species C, there is no behavioral evidence of vocal learning, and no specialized neural implementation is scarce. For species B, there is no behavioral evidence of vocal learning, and no specialized neural implementation for vocal learning is thought to be present (and concomitantly no functional pressures).

<https://doi.org/10.1371/journal.pbio.3000672.g003>

(environment, social feedback [auditory or not], etc.) that help shape the behavior. The “position” of each of these species along each axis can be understood as identifying how much evidence there is for this factor playing a role. Each of the axes, which can be understood as “macrodimensions,” can be further decomposed, and each of the subcomponents would also be subject to different factors.

The choice of “contiguuum” for our conception of vocal learning is intentionally similar to “continuum,” used in the VLC. We do not intend ours to be understood as something completely separate from the VLC but rather as an extension. In this sense, instead of a line, or even a ladder, taking us from category A to B to C and so on, we imagine a morphospace in which different forces bring a species closer to the behavior or pull apart. Like the VLC, we acknowledge the nondichotomous nature of vocal learning abilities. But we go beyond it in considering more factors than VLC for determining such abilities. The VLC would in effect be a vector in our contiguuum: one of several “forces” pushing a species toward one or another phenotype. We name it “specialized neural implementation” in Fig 3. The female zebra finch, for example, which possesses pathways that some considered conducive to vocal learning [34], lacks the behavior readily displayed by males [43], which goes to show that a specific neurobiological pathway cannot be used as a predictor.

Other authors have recently put forward ideas and frameworks that we believe go in a similar direction to ours. Lattenkamp and Vernes [28] and Vernes and Wilkinson [18], though focusing on bats, call attention to the importance of behavioral, developmental, social and

motivational, neurobiological, and ecological factors that play a role in vocal learning. We think this is the right approach to take and that it can be extended to other species: other species can qualify as vocal learners if we accept that neuronal wiring is just one aspect contributing to vocal learning.

These questions of the kind posed by these authors [18, 28], in our view, follow a pedigree of influential work that has ultimately shaped biological research and contributed to a better understanding of cognition. We are referring here specifically to Tinbergen's four questions [85], which ask about mechanism, evolution, ontogeny, and function, and Marr's three levels of analysis [86]: the computational, algorithmic, and implementational levels. These frameworks have forced researchers not to be tied to any one level of description, and keeping all of them in mind when seeking understanding contributes to what Krakauer and colleagues [20] have recently called a pluralistic notion of neuroscience.

Wirthlin and colleagues [26] have a very recent proposal whereby vocal learning can be understood as being made up of different subcomponents, or "modules," and they start by looking at three: vocal coordination (ability to flexibly modify the temporal production of vocal output), vocal production variability (ability to dynamically change acoustic variability throughout development), and vocal versatility (repertoire size versus degree to which it can be modified with experience). Though not exhaustive, these three modules encapsulate several aspects commonly associated with the vocal learning phenotype. Species can be placed along "axes" for each module, and precise comparative and evolutionary characterizations can be attained.

These proposals differ in their details but find commonality in advocating for a multidimensional view of the vocal learning phenotype, which will lead to a more complex but also more accurate representation of its distribution and characteristics. Marrying ideas of this kind with evidence of the kind we review, we believe a more permissive view of vocal learning will start taking shape, encouraging further comparative studies.

## Conclusions

Like other aspects of cognition, vocal learning is a mosaic, made up of different parts. The shared aspects of it should make this even less controversial than, say, language because no one can claim—as they do for language—that what other species have is very different and hard to compare to what we have.

As with any trait, an encompassing view of vocal learning makes it harder to pin down its evolutionary history and the mechanisms behind it. But reducing it to a very specific phenotype and mechanism limits the scope of comparative work, and although it might give the impression that the phenotype becomes more tractable and well-defined, it invariably forces one to subscribe to a very narrow conception that relies on a single driver. Language in general is a good (if extreme) illustration of this. Attempts have been made to reduce language to a very narrow phenotype in order to better study it (e.g., [87]). However, such approaches prevent comparative work almost by definition. As far as we can tell, reductionist views of the language phenotype have not been fruitful and have led to implausible scenarios for the evolution of language [39, 88]. The case of vocal learning is not as extreme because virtually any definition of it yields more than one species with the trait, therefore allowing for some comparative work (although according to the VLC, humans are the only high-end vocal learners). Including more species under the umbrella of any one phenotype should not be a goal in and of itself, but there is much room in our view for casting a wider net and capturing the behaviors that are now being uncovered in other species.

## Supporting information

**S1 Text. Evidence of vocal learning in species not traditionally considered to display vocal learning abilities.**

(PDF)

## References

1. Fitch WT, de Boer B, Mathur N, Ghazanfar AA. Monkey Vocal Tracts Are Speech-Ready. *Science Advances*. 2016; 2(12):e1600723. <https://doi.org/10.1126/sciadv.1600723> PMID: 27957536
2. Tanaka M, Sun F, Li Y, Mooney R. A Mesocortical Dopamine Circuit Enables the Cultural Transmission of Vocal Behaviour. *Nature*. 2018; 563(7729):117–120. <https://doi.org/10.1038/s41586-018-0636-7> PMID: 30333629
3. Jürgens U. The Neural Control of Vocalization in Mammals: A Review. *Journal of Voice*. 2009; 23(1):1–10. <https://doi.org/10.1016/j.jvoice.2007.07.005> PMID: 18207362
4. Mampe B, Friederici AD, Christophe A, Wermke K. Newborns' Cry Melody Is Shaped by Their Native Language. *Current Biology*. 2009; 19(23):1994–1997. <https://doi.org/10.1016/j.cub.2009.09.064> PMID: 19896378
5. Janik VM, Slater PJB. The Different Roles of Social Learning in Vocal Communication. *Animal Behaviour*. 2000; 60(1):1–11. <https://doi.org/10.1006/anbe.2000.1410> PMID: 10924198
6. Shen HH. News Feature: Singing in the Brain. *Proceedings of the National Academy of Sciences*. 2017; 114(36):9490–9493. <https://doi.org/10.1073/pnas.1711799114> PMID: 28874614
7. Samuels BD. Can a Bird Brain Do Phonology? *Frontiers in Psychology*. 2015; 6. <https://doi.org/10.3389/fpsyg.2015.01082> PMID: 26284006
8. Fitch WT. *The Evolution of Language*. Cambridge: Cambridge University Press; 2010.
9. Gill F, Donsker D (eds). *IOC World Bird List (v 9.2)*. 2019; 10.14344/IOC.ML.9.2.
10. Nottebohm F. The Neural Basis of Birdsong. *PLoS Biol*. 2005; 3(5):e164. <https://doi.org/10.1371/journal.pbio.0030164> PMID: 15884976
11. Petkov CI, Jarvis E. Birds, Primates, and Spoken Language Origins: Behavioral Phenotypes and Neurobiological Substrates. *Frontiers in Evolutionary Neuroscience*. 2012; 4. <https://doi.org/10.3389/fnevo.2012.00012> PMID: 22912615
12. Tyack PL, Sayigh LS. Vocal Learning in Cetaceans. In: Snowdon CT, Hausberger M, editors. *Social Influences on Vocal Development*. Cambridge: Cambridge University Press; 1997. p. 208–233.
13. Janik VM. Cetacean Vocal Learning and Communication. *Current Opinion in Neurobiology*. 2014; 28:60–65. <https://doi.org/10.1016/j.conb.2014.06.010> PMID: 25057816
14. Ralls K, Fiorelli P, Gish S. Vocalizations and Vocal Mimicry in Captive Harbor Seals, *Phoca Vitulina*. *Canadian Journal of Zoology*. 1985; 63(5):1050–1056. <https://doi.org/10.1139/z85-157>
15. Ravignani A, Fitch WT, Hanke FD, Heinrich T, Hurgitsch B, Kotz SA, et al. What Pinnipeds Have to Say about Human Speech, Music, and the Evolution of Rhythm. *Frontiers in Neuroscience*. 2016; 10. <https://doi.org/10.3389/fnins.2016.00274> PMID: 27378843
16. Stoeger AS, Manger P. Vocal Learning in Elephants: Neural Bases and Adaptive Context. *Current Opinion in Neurobiology*. 2014; 28:101–107. <https://doi.org/10.1016/j.conb.2014.07.001> PMID: 25062469
17. Prat Y, Taub M, Yovel Y. Vocal Learning in a Social Mammal: Demonstrated by Isolation and Playback Experiments in Bats. *Science Advances*. 2015; 1(2):e1500019. <https://doi.org/10.1126/sciadv.1500019> PMID: 26601149
18. Vernes SC, Wilkinson GS. Behaviour, Biology, and Evolution of Vocal Learning in Bats. *Animal Behaviour and Cognition*. 2019; 375(1789):20190061. <https://doi.org/10.1098/rstb.2019.0061> PMID: 31735153
19. Chakraborty M, Walløe S, Nedergaard S, Fridel EE, Dabelsteen T, Pakkenberg B, et al. Core and Shell Song Systems Unique to the Parrot Brain. *PLoS ONE*. 2015; 10(6):e0118496. <https://doi.org/10.1371/journal.pone.0118496> PMID: 26107173
20. Krakauer JW, Ghazanfar AA, Gomez-Marin A, MacIver MA, Poeppel D. Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*. 2017; 93(3):480–490. <https://doi.org/10.1016/j.neuron.2016.12.041> PMID: 28182904
21. Lieberman P. Vocal Tract Anatomy and the Neural Bases of Talking. *Journal of Phonetics*. 2012; 40(4):608–622. <https://doi.org/10.1016/j.wocn.2012.04.001>



22. Lameira AR. Bidding Evidence for Primate Vocal Learning and the Cultural Substrates for Speech Evolution. *Neuroscience & Biobehavioral Reviews*. 2017; 83(Supplement C):429–439. <https://doi.org/10.1016/j.neubiorev.2017.09.021> PMID: 28947156
23. Tchernichovski O, Marcus G. Vocal Learning beyond Imitation: Mechanisms of Adaptive Vocal Development in Songbirds and Human Infants. *Current Opinion in Neurobiology*. 2014; 28:42–47. <https://doi.org/10.1016/j.conb.2014.06.002> PMID: 25005823
24. Ghazanfar AA, Liao DA, Takahashi DY. Volition and Learning in Primate Vocal Behaviour. *Animal Behaviour*. 2019; <https://doi.org/10.1016/j.anbehav.2019.01.021>
25. Brenowitz EA, Beecher MD. Song Learning in Birds: Diversity and Plasticity, Opportunities and Challenges. *Trends in Neurosciences*. 2005; 28(3):127–132. <https://doi.org/10.1016/j.tins.2005.01.004> PMID: 15749165
26. Wirthlin M, Chang EF, Knörnschild M, Krubitzer LA, Mello CV, Miller CT, et al. A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait. *Neuron*. 2019; 104(1):87–99. <https://doi.org/10.1016/j.neuron.2019.09.036> PMID: 31600518
27. Takahashi DY. Vocal Learning: Shaping by Social Reinforcement. *Current Biology*. 2019; 29(4):R125–R127. <https://doi.org/10.1016/j.cub.2019.01.001> PMID: 30779900
28. Lattenkamp EZ, Vernes SC. Vocal Learning: A Language-Relevant Trait in Need of a Broad Cross-Species Approach. *Current Opinion in Behavioral Sciences*. 2018; 21:209–215. <https://doi.org/10.1016/j.cobeha.2018.04.007>
29. Arriaga G, Zhou EP, Jarvis ED. Of Mice, Birds, and Men: The Mouse Ultrasonic Song System Has Some Features Similar to Humans and Song-Learning Birds. *PLoS ONE*. 2012; 7(10):e46610. <https://doi.org/10.1371/journal.pone.0046610> PMID: 23071596
30. Arriaga G, Jarvis ED. Mouse Vocal Communication System: Are Ultrasounds Learned or Innate? *Brain and language*. 2013; 124(1). <https://doi.org/10.1016/j.bandl.2012.10.002> PMID: 23295209
31. Jarvis ED. Evolution of Vocal Learning and Spoken Language. *Science*. 2019; 366(6461):50–54. <https://doi.org/10.1126/science.aax0287> PMID: 31604300
32. Fitch WT, Zuberbühler K. Primate Precursors to Human Language: Beyond Discontinuity. In: Altenmüller E, Schmidt S, Zimmermann E, editors. *Evolution of Emotional Communication*. Oxford University Press; 2013. p. 26–48.
33. Deacon TW. *The Symbolic Species*. New York: Norton; 1997.
34. Jarvis ED. Learned Birdsong and the Neurobiology of Human Language. *Annals of the New York Academy of Sciences*. 2004; 1016(1):749–777. <https://doi.org/10.1196/annals.1298.038> PMID: 15313804
35. Fitch WT, Huber L, Bugnyar T. Social Cognition and the Evolution of Language: Constructing Cognitive Phylogenies. *Neuron*. 2010; 65(6):795–814. <https://doi.org/10.1016/j.neuron.2010.03.011> PMID: 20346756
36. Kuypers HGJM. Some Projections from the Peri-central Cortex to the Pons and Lower Brain Stem in Monkey and Chimpanzee. *Journal of Comparative Neurology*. 1958; 110(2):221–255. <https://doi.org/10.1002/cne.901100205> PMID: 13654557
37. Jürgens U, von Cramon D. On the Role of the Anterior Cingulate Cortex in Phonation: A Case Report. *Brain and Language*. 1982; 15(2):234–248. [https://doi.org/10.1016/0093-934x\(82\)90058-x](https://doi.org/10.1016/0093-934x(82)90058-x) PMID: 7074343
38. Jürgens U. Neural Pathways Underlying Vocal Control. *Neuroscience & Biobehavioral Reviews*. 2002; 26(2):235–258. [https://doi.org/10.1016/S0149-7634\(01\)00068-9](https://doi.org/10.1016/S0149-7634(01)00068-9)
39. Martins PT, Boeckx C. Language Evolution and Complexity Considerations: The No Half-Merge Fallacy. *PLoS Biol*. 2019; 17(11):e3000389. <https://doi.org/10.1371/journal.pbio.3000389> PMID: 31774810
40. Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, et al. Human-Specific Transcriptional Regulation of CNS Development Genes by FOXP2. *Nature*. 2009; 462(7270):213. <https://doi.org/10.1038/nature08549> PMID: 19907493
41. Katz PS. Evolution of Central Pattern Generators and Rhythmic Behaviours. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2016; 371(1685):20150057. <https://doi.org/10.1098/rstb.2015.0057> PMID: 26598733
42. Lameira AR, Hardus ME, Mielke A, Wich SA, Shumaker RW. Vocal Fold Control beyond the Species-Specific Repertoire in an Orang-Utan. *Scientific Reports*. 2016; 6:30315. <https://doi.org/10.1038/srep30315> PMID: 27461756
43. Shaughnessy DW, Hyson RL, Bertram R, Wu W, Johnson F. Female Zebra Finches Do Not Sing yet Share Neural Pathways Necessary for Singing in Males. *Journal of Comparative Neurology*. 2019; 527(4):843–855. <https://doi.org/10.1002/cne.24569> PMID: 30370534

44. Pidoux L, Le Blanc P, Levenes C, Leblois A. A Subcortical Circuit Linking the Cerebellum to the Basal Ganglia Engaged in Vocal Learning. *eLife*. 2018; 7:e32167. <https://doi.org/10.7554/eLife.32167> PMID: 30044222
45. Kearney MG, Warren TL, Hisey E, Qi J, Mooney R. Discrete Evaluative and Premotor Circuits Enable Vocal Learning in Songbirds. *Neuron*. 2019; 104(3):559–575.e6. <https://doi.org/10.1016/j.neuron.2019.07.025> PMID: 31447169
46. Nottebohm F, Stokes TM, Leonard CM. Central Control of Song in the Canary, *Serinus Canarius*. *Journal of Comparative Neurology*. 1976; 165(4):457–486. <https://doi.org/10.1002/cne.901650405> PMID: 1262540
47. Wild JM, Li D, Eagleton C. Projections of the Dorsomedial Nucleus of the Intercollicular Complex (DM) in Relation to Respiratory-Vocal Nuclei in the Brainstem of Pigeon (*Columba Livia*) and Zebra Finch (*Taeniopygia Guttata*). *Journal of Comparative Neurology*. 1997; 377(3):392–413. [https://doi.org/10.1002/\(sici\)1096-9861\(19970120\)377:3<392::aid-cne7>3.0.co;2-y](https://doi.org/10.1002/(sici)1096-9861(19970120)377:3<392::aid-cne7>3.0.co;2-y) PMID: 8989654
48. Cheney DL, Seyfarth RM. Flexible Usage and Social Function in Primate Vocalizations. *Proceedings of the National Academy of Sciences*. 2018;201717572. <https://doi.org/10.1073/pnas.1717572115> PMID: 29432157
49. Lameira AR, Maddieson I, Zuberbühler K. Primate Feedstock for the Evolution of Consonants. *Trends in Cognitive Sciences*. 2014; 18(2):60–62. <https://doi.org/10.1016/j.tics.2013.10.013> PMID: 24238780
50. Heeren WFL. Vocalic Correlates of Pitch in Whispered versus Normal Speech. *The Journal of the Acoustical Society of America*. 2015; 138(6):3800–3810. <https://doi.org/10.1121/1.4937762> PMID: 26723334
51. Fitch WT. The Biology and Evolution of Speech: A Comparative Analysis. *Annual Review of Linguistics*. 2018; 4(1):255–279. <https://doi.org/10.1146/annurev-linguistics-011817-045748>
52. Okanoya K. Sexual Communication and Domestication May Give Rise to the Signal Complexity Necessary for the Emergence of Language: An Indication from Songbird Studies. *Psychonomic Bulletin & Review*. 2017; 24(1):106–110. <https://doi.org/10.3758/s13423-016-1165-8> PMID: 28120319
53. Takahashi M, Okanoya K. Song Learning in Wild and Domesticated Strains of White-Rumped Munia, *Lonchura Striata*, Compared by Cross-Fostering Procedures: Domestication Increases Song Variability by Decreasing Strain-Specific Bias. *Ethology*. 2010; 116(5):396–405. <https://doi.org/10.1111/j.1439-0310.2010.01761.x>
54. Chakraborty M, Jarvis ED. Brain Evolution by Brain Pathway Duplication. *Phil Trans R Soc B*. 2015; 370(1684):20150056. <https://doi.org/10.1098/rstb.2015.0056> PMID: 26554045
55. Baker A, editor. *The Linguistics of Sign Languages: An Introduction*. Amsterdam; Philadelphia: John Benjamins Pub. Company; 2016.
56. Thomas J, Kirby S. Self Domestication and the Evolution of Language. *Biology & Philosophy*. 2018; 33(1–2). <https://doi.org/10.1007/s10539-018-9612-8> PMID: 29606782
57. Takahashi DY, Liao DA, Ghazanfar AA. Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys. *Current Biology*. 2017; 27(12):1844–1852.e6. <https://doi.org/10.1016/j.cub.2017.05.004> PMID: 28552359
58. Gardner TJ, Naef F, Nottebohm F. Freedom and Rules: The Acquisition and Reprogramming of a Bird's Learned Song. *Science (New York, NY)*. 2005; 308(5724):1046–1049. <https://doi.org/10.1126/science.1108214> PMID: 15890887
59. Fehér O, Wang H, Saar S, Mitra PP, Tchernichovski O. *De Novo* Establishment of Wild-Type Song Culture in the Zebra Finch. *Nature*. 2009; 459(7246):564–568. <https://doi.org/10.1038/nature07994> PMID: 19412161
60. Ferhat AT, Torquet N, Le Sourd AM, de Chaumont F, Olivo-Marin JC, Faure P, et al. Recording Mouse Ultrasonic Vocalizations to Evaluate Social Communication. *Journal of Visualized Experiments*. 2016; 112:53871. <https://doi.org/10.3791/53871> PMID: 27341321
61. Hayes KJ, Hayes C. The Intellectual Development of a Home-Raised Chimpanzee. *Proceedings of the American Philosophical Society*. 1951; 95(2):105–109.
62. Takahashi DY, Fenley AR, Ghazanfar AA. Early Development of Turn-Taking with Parents Shapes Vocal Acoustics in Infant Marmoset Monkeys. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2016; 371(1693):20150370. <https://doi.org/10.1098/rstb.2015.0370> PMID: 27069047
63. Liao DA, Zhang YS, Cai LX, Ghazanfar AA. Internal States and Extrinsic Factors Both Determine Monkey Vocal Production. *Proceedings of the National Academy of Sciences*. 2018; 115(15):3978–3983. <https://doi.org/10.1073/pnas.1722426115> PMID: 29581269
64. Choi JY, Takahashi DY, Ghazanfar AA. Cooperative Vocal Control in Marmoset Monkeys via Vocal Feedback. *Journal of Neurophysiology*. 2015; 114(1):274–283. <https://doi.org/10.1152/jn.00228.2015> PMID: 25925323

65. Miller CT, Wren Thomas A. Individual Recognition during Bouts of Antiphonal Calling in Common Marmosets. *Journal of Comparative Physiology A*. 2012; 198(5):337–346. <https://doi.org/10.1007/s00359-012-0712-7> PMID: 22277952
66. Gultekin YB, Hage SR. Limiting Parental Feedback Disrupts Vocal Development in Marmoset Monkeys. *Nature Communications*. 2017; 8:14046. <https://doi.org/10.1038/ncomms14046> PMID: 28090084
67. Candiotti A, Zuberbühler K, Lemasson A. Convergence and Divergence in Diana Monkey Vocalizations. *Biology Letters*. 2012; 8(3):382–385. <https://doi.org/10.1098/rsbl.2011.1182> PMID: 22237503
68. Lemasson A, Glas L, Barbu S, Lacroix A, Guilloux M, Remeuf K, et al. Youngsters Do Not Pay Attention to Conversational Rules: Is This so for Nonhuman Primates? *Scientific Reports*. 2011; 1:22. <https://doi.org/10.1038/srep00022> PMID: 22355541
69. Lemasson A, Ouattara K, Petit EJ, Zuberbühler K. Social Learning of Vocal Structure in a Nonhuman Primate? *BMC Evolutionary Biology*. 2011; 11:362. <https://doi.org/10.1186/1471-2148-11-362> PMID: 22177339
70. Hage SR, Gavrilov N, Nieder A. Developmental Changes of Cognitive Vocal Control in Monkeys. *The Journal of Experimental Biology*. 2016; 219(11):1744–1749. <https://doi.org/10.1242/jeb.137653> PMID: 27252457
71. Lameira AR, Hardus ME, Bartlett AM, Shumaker RW, Wich SA, Menken SBJ. Speech-Like Rhythm in a Voiced and Voiceless Orangutan Call. *PLoS ONE*. 2015; 10(1):e116136. <https://doi.org/10.1371/journal.pone.0116136> PMID: 25569211
72. Wich SA, Swartz KB, Hardus ME, Lameira AR, Stromberg E, Shumaker RW. A Case of Spontaneous Acquisition of a Human Sound by an Orangutan. *Primates*. 2009; 50(1):56–64. <https://doi.org/10.1007/s10329-008-0117-y> PMID: 19052691
73. Lameira AR, Hardus ME, Kowalsky B, de Vries H, Spruijt BM, Sterck EHM, et al. Orangutan (*Pongo* Spp.) Whistling and Implications for the Emergence of an Open-Ended Call Repertoire: A Replication and Extension. *The Journal of the Acoustical Society of America*. 2013; 134(3):2326–2335. <https://doi.org/10.1121/1.4817929> PMID: 23967963
74. Hardus ME, Lameira AR, Schaik CPV, Wich SA. Tool Use in Wild Orang-Utans Modifies Sound Production: A Functionally Deceptive Innovation? *Proceedings of the Royal Society of London B: Biological Sciences*. 2009; 276(1673):3689–3694. <https://doi.org/10.1098/rspb.2009.1027> PMID: 19656794
75. Lameira AR, Hardus ME, Wich SA. Orangutan Instrumental Gesture-Calls: Reconciling Acoustic and Gestural Speech Evolution Models. *Evolutionary Biology*. 2012; 39(3):415–418. <https://doi.org/10.1007/s11692-011-9151-6> PMID: 22923853
76. Perlman M, Clark N. Learned Vocal and Breathing Behavior in an Enculturated Gorilla. *Animal Cognition*. 2015; 18(5):1165–1179. <https://doi.org/10.1007/s10071-015-0889-6> PMID: 26139343
77. Watson SK, Townsend SW, Schel AM, Wilke C, Wallace EK, Cheng L, et al. Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. *Current Biology*. 2015; 25(4):495–499. <https://doi.org/10.1016/j.cub.2014.12.032> PMID: 25660548
78. Holy TE, Guo Z. Ultrasonic Songs of Male Mice. *PLoS Biol*. 2005; 3(12):e386. <https://doi.org/10.1371/journal.pbio.0030386> PMID: 16248680
79. Grimsley JMS, Monaghan JJM, Wenstrup JJ. Development of Social Vocalizations in Mice. *PLoS ONE*. 2011; 6(3):e17460. <https://doi.org/10.1371/journal.pone.0017460> PMID: 21408007
80. Miller JR, Engstrom MD. Vocal Stereotypy and Singing Behavior in Baiomyine Mice. *Journal of Mammalogy*. 2007; 88(6):1447–1465. <https://doi.org/10.1644/06-MAMM-A-386R.1>
81. Okobi DE, Banerjee A, Matheson AMM, Phelps SM, Long MA. Motor Cortical Control of Vocal Interaction in Neotropical Singing Mice. *Science*. 2019; 363(6430):983–988. <https://doi.org/10.1126/science.aau9480> PMID: 30819963
82. Chamberlain SA, Szöcs E. Taxize: Taxonomic Search and Retrieval in R. *F1000Research*. 2013; 2:191. <https://doi.org/10.12688/f1000research.2-191.v2> PMID: 24555091
83. Martins PT, Boeckx C. Sound Production Learning and the Contiguity Hypothesis. *Proceedings of the 12th International Conference on the Evolution of Language (Evolang12)*. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika; 2018. p. 284–286.
84. Theofanopoulou C, Boeckx C. Cognitive Phylogenies, the Darwinian Logic of Descent, and the Inadequacy of Cladistic Thinking. *Frontiers in Cell and Developmental Biology*. 2015; 3. <https://doi.org/10.3389/fcell.2015.00064> PMID: 26528479
85. Tinbergen N. On Aims and Methods of Ethology. *Zeitschrift für Tierpsychologie*. 1963; 20(4):410–433.
86. Marr D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W.H. Freeman; 1982.
87. Berwick RC, Chomsky N. *Why Only Us; Language and Evolution*. Cambridge, MA: MIT Press; 2016.

88. de Boer B, Thompson B, Ravnani A, Boeckx C. Evolutionary Dynamics Do Not Motivate a Single-Mutant Theory of Human Language. *Scientific Reports*. 2020; 10(1):451. <https://doi.org/10.1038/s41598-019-57235-8> PMID: 31949223