



Universiteit  
Leiden  
The Netherlands

## **The multi-dimensional nature of vocal learning**

Vernes, S.C.; Kriengwatana, B.P.; Beeck, V.C.; Fischer, J.; Tyack, P.L.; Cate, C.J. ten; Janik, V.M.

### **Citation**

Vernes, S. C., Kriengwatana, B. P., Beeck, V. C., Fischer, J., Tyack, P. L., Cate, C. J. ten, & Janik, V. M. (2021). The multi-dimensional nature of vocal learning. *Philosophical Transactions Of The Royal Society Of London Series B: Biological Sciences*, 376(1836). doi:10.1098/rstb.2020.0236

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3254608>

**Note:** To cite this publication please use the final published version (if applicable).

Review



**Cite this article:** Vernes SC, Kriengwatana BP, Beeck VC, Fischer J, Tyack PL, ten Cate C, Janik VM. 2021 The multi-dimensional nature of vocal learning. *Phil. Trans. R. Soc. B* **376**: 20200236.

<https://doi.org/10.1098/rstb.2020.0236>

Accepted: 26 May 2021

One contribution of 21 to a theme issue ‘Vocal learning in animals and humans’.

**Subject Areas:**

behaviour, cognition, ecology

**Keywords:**

vocal learning, songbird, cognition, behaviour, evolution, language

**Authors for correspondence:**

Sonja C. Vernes

e-mail: [scv1@st-andrews.ac.uk](mailto:scv1@st-andrews.ac.uk)

Vincent M. Janik

e-mail: [vj@st-andrews.ac.uk](mailto:vj@st-andrews.ac.uk)

# The multi-dimensional nature of vocal learning

Sonja C. Vernes<sup>1,2,3</sup>, Buddhadas Pralle Kriengwatana<sup>4</sup>, Veronika C. Beeck<sup>5</sup>, Julia Fischer<sup>6,7</sup>, Peter L. Tyack<sup>1</sup>, Carel ten Cate<sup>8</sup> and Vincent M. Janik<sup>1</sup>

<sup>1</sup>School of Biology, University of St Andrews, St Andrews, UK

<sup>2</sup>Neurogenetics of Vocal Communication Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

<sup>3</sup>Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands

<sup>4</sup>Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK

<sup>5</sup>Department of Behavioural and Cognitive Biology, University of Vienna, Vienna, Austria

<sup>6</sup>Cognitive Ethology Laboratory, German Primate Centre, Göttingen, Germany

<sup>7</sup>Department of Primate Cognition, Georg-August-University Göttingen, Göttingen, Germany

<sup>8</sup>Institute of Biology, Leiden University, Leiden, The Netherlands

SCV, 0000-0003-0305-4584; JF, 0000-0002-5807-0074; PLT, 0000-0002-8409-4790; CtC, 0000-0002-4021-8915; VMJ, 0000-0001-7894-0121

How learning affects vocalizations is a key question in the study of animal communication and human language. Parallel efforts in birds and humans have taught us much about how vocal learning works on a behavioural and neurobiological level. Subsequent efforts have revealed a variety of cases among mammals in which experience also has a major influence on vocal repertoires. Janik and Slater (*Anim. Behav.* **60**, 1–11. (doi:10.1006/anbe.2000.1410)) introduced the distinction between vocal usage and production learning, providing a general framework to categorize how different types of learning influence vocalizations. This idea was built on by Petkov and Jarvis (*Front. Evol. Neurosci.* **4**, 12. (doi:10.3389/fnevo.2012.00012)) to emphasize a more continuous distribution between limited and more complex vocal production learners. Yet, with more studies providing empirical data, the limits of the initial frameworks become apparent. We build on these frameworks to refine the categorization of vocal learning in light of advances made since their publication and widespread agreement that vocal learning is not a binary trait. We propose a novel classification system, based on the definitions by Janik and Slater, that deconstructs vocal learning into key dimensions to aid in understanding the mechanisms involved in this complex behaviour. We consider how vocalizations can change without learning, and a usage learning framework that considers context specificity and timing. We identify dimensions of vocal production learning, including the copying of auditory models (convergence/divergence on model sounds, accuracy of copying), the degree of change (type and breadth of learning) and timing (when learning takes place, the length of time it takes and how long it is retained). We consider grey areas of classification and current mechanistic understanding of these behaviours. Our framework identifies research needs and will help to inform neurobiological and evolutionary studies endeavouring to uncover the multi-dimensional nature of vocal learning.

This article is part of the theme issue ‘Vocal learning in animals and humans’.

## 1. Introduction

Interest in vocal learning skills has been with us from the beginning of animal behaviour research, with scholars as early as Darwin [1] recognizing the role of learning in the development of bird song and its parallels with human vocal

performance. Janik & Slater [2,3] introduced a framework that distinguished between vocal usage learning, in which existing signals are given in a new context or sequence, and vocal production learning, in which signals are modified in form after experience with the signals of others. Petkov & Jarvis [4] noted the continuous nature of this trait, and others have further explored these ideas (e.g. [5–10]). Yet, with more studies providing empirical data, the limits of the initial frameworks become apparent. We build on the definitions proposed by Janik & Slater [2] to refine the categorization of vocal learning in light of the advances made since its publication. Our motivation is that, despite the advances made in the last 20 years and more of research since the first publication of this framework, identifying which species are vocal learners still remains a challenging task. Outside the few highly studied and easily recognizable vocal production learners, membership of this select club is still hotly debated. Furthermore, despite most researchers agreeing that vocal learning is not a binary trait, a satisfactory typology of its components is still lacking.

Vocalizations may vary for a range of reasons, but vocal learning crucially requires a learning process to be the driver of such variation. Although many definitions of learning exist (e.g. see [11]), we have adopted the definition of learning as a modification of an individual's behaviour owing to information memorized from previous experience [12]. Vocal production learning was defined by Janik & Slater [2] as the production of modified or novel vocalizations, as a result of learning from the experience of the acoustic signals of others. Central to this definition is that auditory input from a model sound leads to the formation of a memory of the sound (a template), to which the vocal output is compared. We maintain this definition and highlight that the auditory model may be a range of sounds including vocalizations produced by another animal (recorded or live), synthesized playbacks or even non-vocal mechanical sounds (e.g. lyrebirds' imitation of a chainsaw or camera shutter).

We classify behavioural changes associated with vocal learning into a cluster of discrete dimensions and then consider how these behaviours relate to evidence for or against underlying mechanisms. This approach explicitly acknowledges that there may not be a single vocal learning phenotype, rather multiple components, expressed to different degrees. We first consider cases where changes in vocalizations may be occurring that are not the result of learning, to distinguish them from learned modifications (§2). We then explore the definitions of, and distinctions between, vocal usage learning and vocal production learning. Usage learning is a form of contextual learning in which a signal is produced in a new context as a result of experience [2]. This term covers a range of behaviours where the context may refer to environmental or behavioural contexts (natural or artificial), and can include positional contexts (i.e. position in a sequence of calls). For usage learning (§3), we consider two dimensions: how vocalization context (behavioural contexts and positional context) or timing (call timing and rhythmicity) is learned. For vocal production learning itself (§4), we describe three dimensions: the copying of auditory models (convergence or divergence on model sounds, accuracy of copying), the degree of change (type of vocal modification and breadth of learning) and timing (when learning takes place, the length of time it takes and how long it is retained). We also consider the grey areas that currently hinder clear classification (§5), our

current mechanistic understanding of these traits (§6) and outstanding questions for future research (§7 and Box 1).

Our framework gives us a means to directly compare the different aspects of vocal learning abilities across animals. In this way, we aim to make clear the research directions that are needed to close current gaps in knowledge and make significant strides in understanding the remarkable trait of vocal learning. We hope this framework will make it possible to investigate and reveal the mechanisms that drive each dimension, both in individual species, or in a true 'like-with-like' cross-species approach, and in this way, better understand the prevalence and evolution of this complex phenomenon.

## 2. Non-learned inputs into vocal variation

To make headway in the identification of vocal learning, it is necessary to first recognize other factors that may contribute to variation, but that do not involve the learned adjustment of vocal output.

### (a) Variation in response to the acoustic environment

Animals modify the intensity, frequency, duration and repetition of their signals to compensate for variation in acoustic conditions in their environment (see [13–16]). If these adjustments involve immediate audio-vocal feedback [17–19] rather than memorized information from previous experience, they would not be classed as learned changes, based on the definition of learning applied herein. This includes phenomena such as the jamming avoidance response of some bat species [20].

Highly sophisticated or precise vocal change in response to auditory input does not necessarily indicate that the change is learned. One sophisticated phenomenon that involves auditory–vocal feedback but is not generally defined as vocal learning is the Doppler shift compensation (DSC) of Rhinolophid bats. When a bat hears a Doppler-shifted echo, it precisely modifies the frequency of its call so that the echo stays in a preferred frequency band [21]. DSC is achieved by auditory–vocal feedback mechanisms in the midbrain that do not require learning under our definition [22]. Thus, to classify observed vocal variation in response to environmental sounds, it is critical to determine if the change is a non-learned response (such as DSC), or whether the animal employs a learning process to modify its vocal output.

### (b) Controlling for other sources of acoustic variation

Changes owing to physical growth and maturation (e.g. [23]), rather than changes in auditory experience, can cause a shift in vocal characteristics (e.g. fallow deer (*Dama dama*), [24]). Changes in arousal, stress, disease and hormonal or emotional state (e.g. [25,26]) can also account for changes in acoustic structure [27]. For example, the rate of parental vocal input was shown to positively affect the rate of vocal maturation in common marmoset young [28,29]. These differences may be explained by an increased effort in practising driving an existing developmental programme more rapidly. Stress, e.g. induced by separation, can have a direct impact on the spectral structure of calls in common marmosets [30] and the rate of parental vocal input can have a stress alleviating effect. Acoustic features of calls may also be influenced by the reproductive state of the subjects [31]. Therefore, the

contribution of physiological and morphological factors to changes in vocalizations should be considered before inferring vocal learning.

### 3. Vocal usage learning

Usage learning and vocal production learning are two different facets of how learning can affect vocal behaviour. Usage learning can be applied to existing vocalizations or those acquired by vocal production learning, and some animals are capable of both. So, while usage and production learning can co-occur, usage learning does not itself require vocal production learning. For instance, if a young animal learns to restrict the use of an alarm call to a specific set of predators, but the alarm call itself does not require learning to develop, this is a clear example of usage learning. However, if an animal learns to produce a novel vocalization to signal the presence of predators this would involve vocal production learning (to learn the novel vocalization) and usage learning (to learn the context in which to employ it).

Vocal usage learning appears to be fairly common in mammals and birds. Adret [32] and Janik & Slater [2] highlighted usage learning in a broad group of species including parrots, chickens, budgerigars, mynahs, cats, dogs, rats, lemurs, sea lions and dolphins. More recent studies add bats [33], elephants [34] and seals [35] to this list. Nevertheless, usage learning can take various different forms and levels of complexity and describing these quite diverse phenomena under the label of usage learning does not imply that a single mechanism or the same one underlies them all. For example, primatologists have proposed that non-human primates have a special propensity for usage learning that is a precursor to learning in human language development [9,36]. Below we describe some of the different contexts in which the use of signals can change through usage learning.

#### (a) Learning to use calls in a new context

The most widely recognized form of usage learning is learning to produce an existing call in a new environmental context. A clear example of this is a study on European blackbirds (*Turdus merula*). Perceiving a mobbing conspecific together with a novel, harmless stuffed bird (a honey eater) induced blackbirds to mob the innocuous object using specific mobbing vocal behaviour (e.g. 'duck' calls) normally given to predators. This behaviour could be culturally transmitted along a chain of at least six individuals [37].

A common way of demonstrating learning to use calls in a new context is by training an animal to produce vocalizations in response to a conditioned stimulus [2] such as an arbitrary hand or light signal, which can be considered a fairly simple form of contextual usage learning. Studies on a variety of species used such a method to demonstrate conditioned control over a single call type [32]. More sophisticated control over vocal usage was shown by crows that were able to produce or withhold vocalizing to different colour cues [38]. Rhesus monkeys (*Macaca mulatta*) [39], gray seals (*Halichoerus grypus*) [35] and bottlenose dolphins (*Tursiops truncatus*) [40] were even able to learn to produce more than one call type in response to different arbitrary signals. Note that, in the case of the dolphins, these were novel synthetic calls that were learned and used in an arbitrary context [40], suggesting a case where both vocal production and usage learning were

occurring. Another method is to link food delivery to the production of specific signal parameters without training animals to respond to specific signals. Pale spear-nosed bats (*Phyllostomus discolor*) shifted to rewarded versions of their social calls in such a setup [33].

Identifying usage learning in the wild is more challenging. Animals often change how they use calls, as well as which calls they use over the course of their development but as before it is difficult to distinguish changes caused by maturation or stress from those caused by learning. Natural experiments can help here when populations differ in the number of examples provided to a developing animal. For example, groups of vervet monkeys (*Cercopithecus aethiops*) with more intergroup encounters learned the correct use of 'wrr' calls more quickly than groups with fewer encounters [41]. Another interesting context in which usage learning might occur is the deceptive use of calls. For example, several species occasionally produce alarm calls in the absence of any predators, often gaining access to food in the process (e.g. [42]). If shown not to be a production error, the development of this behaviour is likely to involve usage learning. Fork-tailed drongos (*Dicrurus adsimilis*) also produce alarm calls of different species depending on what species are in their audience [43]. Drongos use production learning to acquire the alarm calls of other species, and usage learning to learn the context in which they are given. Developmental studies of these behaviours will be able to shed further light on the role of usage learning in such contexts.

Animals may also learn to produce calls in a new positional context or sequence [2,5]. This can involve the sequencing of calls from an existing repertoire or the sequencing of novel calls acquired through vocal production learning. The arrangement of calls can be in simple sequences, or can apply complex syntactic rules. Arranging or rearranging the sequence of calls or song elements might occur in various ways and through various mechanisms. Depending on the process involved in specific cases, such sequencing may be classified as usage learning or as production learning, or may not fit neatly into these categories. For this reason, we address sequence learning in §5 and acknowledge we need a deeper understanding of the mechanisms underlying sequence learning in the future.

#### (b) Learning timing of call use: vocal interactions and rhythms

Coordination of call timing is widespread across the animal kingdom, from coordination of utterances in human conversations [44], to temporally entrained advertisement calls of male frogs [45]. Duetting may involve temporal coordination such that vocalizations overlap, or occur in a turn-taking pattern with calls and their responses being produced within a limited time window. The widespread nature of this ability has suggested it as relevant for the evolution of communication in animals and humans [46]. However, turn taking does not necessarily require learning and it is important to make a distinction between those temporal entrainments that are learned, those that are not learned, and those for which learning has not yet been demonstrated.

Temporal coordination of calls often develops very early, making it challenging to determine if the behaviour is learned or not, but examples where call timing influenced by learning is particularly apparent are in the development



of duetting displays in birds [47] and mammals [48]. Some bird species produce simple call and response vocalizations, while others produce highly sophisticated duets that are so precisely timed that it can be difficult to tell when one bird stops singing and another continues the song (e.g. [49]). In canebrake wrens (*Cantorchilus zeledoni*), juveniles start duetting by singing together with adults, and their coordination of songs becomes better over time [50]. Furthermore, adults that acquire a new partner have poor coordination directly after pairing, but this improves the longer they duet together [47]. The same has been found in gibbons [48,51]. These findings suggest that canebrake wrens and gibbons learn the appropriate temporal coordination for duetting.

Human and non-human primates (e.g. common marmosets) demonstrate vocal turn-taking behaviour from early in life [28,44,52,53]. Many other mammals such as cetaceans [54], bats [55], elephants [56,57], meerkats [58] and naked mole rats [59] also show turn-taking behaviour, although in most cases learning of timing has not been definitively demonstrated in mammals. By contrast, there are some vertebrates where temporal vocal coordination can clearly be attributed to a central nervous system oscillator that is responsive to call perception, rather than a learned mechanism, such as in rhythmic calls of some anurans [45] or suppression of motor pattern generators in response to hearing conspecific calls in monkeys [60].

Another form of usage learning found in human speech and song is copying novel rhythms in vocalization patterns. Learning to produce novel vocal rhythms has not been demonstrated in animals even though examples of entrainment of motor responses to acoustic rhythms exist [61,62]. Zebra finches (*Taeniopygia guttata*) can create a type of rhythm by predicting the timing of a regular jamming signal and adjusting their own vocalizations to avoid it [63]. They can also learn to distinguish perceptually between a specific isochronous and irregular beat pattern, but do not generalize this to tempo changes of the same patterns, suggesting that they do not have a representation of the global rhythm [64]. Sperm whales (*Physeter macrocephalus*) use rhythmic patterns of clicks, called codas, for communication [65]. Codas are shared across multiple groups, with multiple sympatric vocal clans in the Pacific and Atlantic Oceans [66]. This distribution of codas suggests that whales learn group-distinctive calls through memorizing the rhythmic patterns of their click sounds. Sperm whales are able to precisely time their clicks to match the depth sounder on a ship [67], but learning of rhythmic patterns like those of codas has not been demonstrated. Birds and mammals clearly pay attention to aspects of rhythm in conspecific vocalizations (e.g. [68,69]) but production studies demonstrating vocal learning of rhythm in animals are still missing.

### (c) Constraints and flexibility of usage learning

Training animals to produce their own vocalizations on command is a common way of demonstrating usage learning. As noted above, many species have demonstrated this ability, but few studies have explored how flexible this form of learning is. For example, many species have been trained to produce a signal from their own repertoire on command, but it is unknown if they could be trained to use all of their existing call types in this fashion, or only some of them. The level of difficulty of usage learning in such studies is often reflected in the

time required to learn to produce existing calls in response to different signals [70]. In relation to sequence learning, it is also an open question whether animals are able to re-arrange all elements in a sequence arbitrarily or whether there are constraints. There are often predispositions in bird song learning for conspecific song elements and structure [71,72]. It is, therefore, likely that there are limits to how units are recombined since models with unusual syllable combinations may not fit the species template for the conspecific song. More studies are needed to explore such limitations systematically.

A good example of complexity in usage learning is the fine tuning of the contexts in which vervet monkeys give alarm calls during their development [73]. While they seem to be predisposed from birth to produce an aerial alarm to threats in the sky, the attention to detail required to learn the correct predator species for the call moving from broader categories to more specific ones is considerable. This probably relies on different neural processing than that involved in learning to produce an unspecified vocalization on command in a conditioning test. The most detailed documentation of usage learning complexity comes from language studies with grey parrots (*Psittacus erithacus*) that had to harness usage learning to use human words (which were initially acquired by production learning) to refer to objects and give information about their properties [74]. While speech learning in grey parrots is impressive, their usage learning abilities are not unbounded. For example, the acquisition of multiple labels for the same object is much more difficult than the learning of the initial label [75]. All of these examples are case studies in which usage learning was part of a larger investigation of cognitive skills.

To conclude, it is clear that the term ‘usage learning’ covers a broad category of phenomena, most likely involving a range of different mechanisms. A more systematic investigation of usage learning, its diversity, complexity and its constraints would be of great value.

## 4. Vocal production learning

The classical view of vocal production learning is based on studies of humans and songbirds. In this process, the animal uses auditory input from a model sound to produce copies of various degrees of fidelity, depending on the species. Producing a copy of an auditory model is also known from other animal groups, such as parrots [76], pinnipeds [77] and dolphins [40]. These represent clear cases of vocal production learning, where it can be directly shown that individuals exposed to different model sounds start selectively producing what they have heard. In most songbirds, the neural representation of the model is built up from several exposures and stored as the ‘template’ on which later vocal production is modelled. However, there are also examples in which the copying of a model is almost instantaneous, suggestive of an ability for real-time vocal matching. It is not always clear whether this results in a longer lasting or a perhaps more transient neural representation of the auditory model.

Our definition of vocal production learning requires auditory input. However, it is important to note that this does not imply that auditory cues are the only factors involved in this learning process. It has been clearly shown that other factors, such as social interactions, can influence how effectively animals learn during vocal production learning. In humans and

songbirds, learning is more effective when interaction with a live tutor is possible, compared to auditory only or even audio-visual playback on a screen [78–80]. As such our definition denotes only the requirement for exposure to an auditory model to form a template during vocal production learning, and should not be interpreted as an exclusion of other contextual factors that clearly play an important role in this process.

Starting from the aforementioned definition, we consider vocal production learning to be an ‘umbrella’ term covering a number of behavioural dimensions, many of which appear to be continuous, while others are discrete (figure 1). Below, we break down vocal production learning into dimensions to produce a more fine-grained view of this complex behaviour and to show how these dimensions can alter vocal outputs in different ways. In doing so, our aim is to produce a behaviour-focused framework within which it would be possible to categorize the varied vocal production learning abilities of different animals, facilitate inter-species comparisons of these behaviours and ultimately, elucidate the underlying mechanisms.

### (a) MODEL: vocalizations converge on or diverge from the model

The examples of vocal production learning given above result in learning to make vocalizations more similar to an auditory model. However, vocal production learning can also involve modification of vocal outputs so that they become different from the auditory model. Diverging from the model can be beneficial if the vocal learning serves to produce a signal necessary for individual recognition that should be species typical, yet distinct from others that they encounter [2,3]. Bottlenose dolphins use vocal learning in the development of individually distinctive signature whistles seemingly by modifying whistles that they have heard in their environment [92]. Rapid call convergence and divergence has been observed in playback experiments with several parrot species. The degree and direction of the resulting changes seems to depend on the social dynamics of the species and individuals concerned [93]. To diverge from a model, a template based on the model would still form, but the vocal outputs would be modified to avoid similarity to the template. It is possible that a second auditory template could form if reliable production of the divergent vocalization is needed. In practice, it may be difficult to determine whether an observed divergent vocalization is the result of active modification to differ from a model, or matching to another model that is unknown to the observer. While convergence on a model is the most widely (and easily) observed behaviour, further work is needed to determine the spread of, and mechanisms underlying, divergence from a model.

### (b) MODEL: accuracy of the copy

The fidelity of a copy produced during vocal production learning has previously been well laid out in relation to song learning in birds by Beecher & Brenowitz [6]. They describe a continuum from high fidelity (faithful imitation), to medium fidelity (improvisation that introduces variation to sounds copied from a tutor), to low fidelity (invention that is species typical but does not resemble a tutor, and may not even require a tutor). In this latter case, the bird may still require exposure to songs in order to produce a species-typical song. For instance, grasshopper sparrows (*Ammodramus savannarum*) must have

heard normal songs in order to develop normal, species-typical warble songs instead of impoverished songs with abnormal warble notes [94]. Nevertheless, the songs resulting from exposure to a tutor do not resemble the tutor song more than they resemble another normal grasshopper sparrow song. So, there is an effect of experience, but it does not translate into copying. In a species that produces excellent copies of model sounds heard earlier a clear template must have been formed. In a species requiring exposure to song, but not resulting in copying, the exposure may be needed to stimulate the development of a latent song programme or to learn the appropriate characteristics of species-typical songs. These different outcomes of exposure suggest different underlying processes by which experience can influence vocal development and these contrasting phenomena need further exploration.

There may be circumstances where copying fidelity is of more or less importance during vocal production learning. For example, learning that results in convergence towards a model (see the previous section) may require medium to high fidelity because accuracy brings benefits to the copier. An adaptive need to diverge from a model may employ precise avoidance of the model, or could employ low fidelity copying in combination with improvisation or innovation. For example, bottlenose dolphins address each other by copying their individually distinctive signature whistles [95,96]. While the overall modulation pattern of model whistles is produced accurately in these copies, multi-dimensional scaling of specific acoustic parameters of copies often result in substantial differences from the whistle that is being copied, suggesting a possible use of such mechanisms to make copies recognizable as such to eavesdroppers [97]. In some species of songbird, the ability to accurately imitate song elements or sequences of song elements may be targets of inter- and intra-sexual selection because it could signal aspects of individual quality and fitness [98–101], including motivation and/or ability to attend to social cues, use social information and form social bonds [102,103].

Importantly, when assessing the precision of the match and the importance of fidelity, we must strive to consider this from the animal’s perspective. Although we may be able to quantify subtle differences on a spectrogram, differences in the fidelity of a copy are only relevant when and if the study species can perceive a difference. Conversely, using our own assessment of a ‘good’ match may lead to overlooking important features that are relevant for the study species.

### (c) DEGREE: type of vocal modification

Production learning may involve modifications of existing vocalizations or the production of completely novel vocalizations. In addition, vocal outputs can be modified by learning individual calls or learning sequences of calls (see §5 for further discussion of this issue). In all these cases, a key challenge is to determine what constitutes a novel sound pattern. In determining whether the vocalization observed is a novel sound, or a modification of an existing sound, the complete vocal repertoire for an individual is needed. While this requires knowledge of an individual’s repertoire before it has been exposed to the learning experience, this knowledge can seldom be claimed to be complete because it is very difficult to be sure that all rarely used vocalizations or heavily context-specific vocalizations have been captured. However, in practice, larger qualitative or quantitative changes in

dimensions of vocal production learning		zebra finch	African grey parrot	lyrebird	Costa's hummingbird	Horsfield's bronze cuckoo	bottlenose dolphin	gray seal	spear nosed bat	Asian elephant	? = unknown N/A = not applicable
	auditory input leads to formation of a memory to which vocal output is compared	[82]	[74]	[107]	[84]	[123]	[40]	[77]	[86]	[34]	no <input type="radio"/> <input checked="" type="radio"/> yes
model	vocalisations converge on or diverge from the model	[82]	[74]	[107]	[84]	N/A	[92]	[77]	[86]	[34]	diverge <input type="radio"/> either <input type="radio"/> converge <input checked="" type="radio"/>
	accuracy of the copy	[82]	[74]	[107]	[84]	N/A	[40]	[77]	[121]	[34]	poor <input type="radio"/> <input checked="" type="radio"/> good
degree	type of vocal modification	[82]	[74]	[107]	[84]	[123]	[40]	[77]	[86]	[34]	modify existing <input type="radio"/> <input checked="" type="radio"/> learn new
	breadth of learning	[82]	[74]	[107]	[84]	[123]	[40]	[77]	[86]	[34]	species specific <input type="radio"/> <input checked="" type="radio"/> broad mimicry
timing	when learning takes place	[82]	[74]	?	[88]	?	[40] [89]	?	[121]	?	limited period <input type="radio"/> <input checked="" type="radio"/> through-out life
	length of time learning takes	[82]	[74]	?	[84]	[123]	[40]	[77]	[121]	?	slow <input type="radio"/> <input checked="" type="radio"/> fast
	how long the learning is retained	[82]	[122]	?	[84]	?	[91]	?	?	[34]	briefly <input type="radio"/> <input checked="" type="radio"/> long term

**Figure 1.** The multi-dimensional nature of vocal production learning. Vocal production learning is the production of modified or novel vocalizations, as a result of learning from the experience of the acoustic signals of others. In this process, auditory input leads to the formation of a memory, to which the vocal output is compared. Examples of avian and mammalian vocal production learning species are represented across the dimensions of vocal production learning outlined in §4. The Horsfield's bronze cuckoo is also included based on our discussion of its inclusion in the wider definition of vocal production learning (see §5). This is not an exhaustive list, rather a selection of species with shared and contrasting abilities. Categorizations are made for each species using discrete or continuous scales, as indicated in the legend. This schema can be used to take an in depth look at the vocal learning properties of an individual species (by looking down a column) or to compare abilities within a single dimension of vocal learning (by looking along a row). Categorizations are intended to indicate what each species has been observed to be capable of doing, and may not reflect their behaviour in their normal environment, or the full extent of their capacity. For example, Asian elephants have been demonstrated to have a high breadth of learning largely based on an Asian elephant in human care that learned human speech-like sounds [34]. However, this behaviour is not observed in normally reared Asian elephants in the wild. African elephants in human care are reported to imitate the sounds of Asian elephants and of trucks [81], but the breadth of mimicry for copying other sounds in elephants is not known. A question mark (?) indicates where an accurate categorization cannot be made based on current published work and further research is needed. In the case of the Horsfield's bronze cuckoo, learning does not appear to employ auditory input of a model sound and as such, the other categories within that dimension are labelled not applicable (N/A). All classifications are made based on current available knowledge (see reference numbers within circles for evidence used for classifications), and should be updated as new information comes to light.

vocal structure, in particular, when they are towards some identifiable model, will be strong indicators of learning. In the cases of animals imitating human speech or a songbird copying a mechanical sound, it is obvious that a novel vocalization is being produced. It is not yet known whether modification of existing calls and production of novel calls employ the same or different neural mechanisms [5,17], but this is a key area for future research, as understanding the mechanisms will be highly beneficial not only to better categorize the behaviour, but to understand how the behaviour arises.

Modification of calls can employ different physical mechanisms operating on the airflow, the sound-producing organ or the filtering vocal tract that can co-occur or to some extent independently affect a call's spectral structure. While duration and amplitude need control of the respiratory system this needs to be coordinated with the configuration of the sound source. Similarly, frequency modulation patterns are most prominently influenced by the phonatory system [2] but increasing amplitude by controlling respiratory pressure can lead to an increase in frequency parameters and occurrences of nonlinear phenomena, again affecting the spectral structure. Another important way of learning to modify sounds is by changing the way in which the upper vocal tract filters or emphasizes parts of the signal produced by the sound source (e.g. the vocal folds in most mammals or

syrinx in birds) [104]. Janik & Slater [2] called this filter learning and highlighted its possible involvement in human speech owing to the importance of these structures to encode contrasts in vowels and consonants [105].

#### (d) DEGREE: breadth of learning

Vocal learners differ greatly in the breadth or narrowness of the range of vocalizations they can learn to produce. Several songbird species are strongly biased towards learning to produce only species-typical vocalizations [72]. Some animal species can mimic species-atypical sounds such as human speech (e.g. parrots [74], elephants [34], seals [77]) and some bird species are able to mimic a huge range of sounds produced by different animals (e.g. mynah birds, starlings, drongos [106,83]), or even non-biological sources in their environment. For example, lyrebirds can copy the sounds of a range of other animals or even a chainsaw or camera shutter from their environment [107,108]. Beecher & Brenowitz [6] refer to this variation in breadth of learning as the degree of canalization.

The breadth or narrowness of the vocalizations an animal can learn to produce may result from different selection pressures leading to a more or less narrow choice of models and vocalizations. Morphology may also preclude species from producing certain types of sounds because of the limitations of their vocalizing organ. Neural constraints may



determine whether or not the animal recognizes sounds as a relevant model to attempt to copy and whether motor programmes are flexible enough to copy atypical models.

Environmental factors or motivation may also be key influences in determining the breadth of learning for many species. The social relationship between tutor and tutee and the presentation mode (live or playback) may affect the motivation to copy vocalizations [102]. Similarly, under normal circumstances, and in the normal environment of the animal, it may be the case that only species-typical sounds are learned. However, under unusual conditions (e.g. captive rearing, atypical social bonding with heterospecifics), animals may mimic unusual sounds. Such a situation is exemplified by the individual cases of Hoover the harbour seal (*Phoca vitulina*) [109], Koshik the Indian elephant (*Elephas maximus*) [34] and most recently by a case of Ripper, an Australian musk duck (*Biziura lobata*) [110], which were each observed to mimic human speech-like sounds. Seals, elephants and musk ducks are not routinely observed to mimic human speech, but in each of these cases, the individuals were reared in isolation from conspecifics and formed strong bonds with human keepers. It is hypothesized that in the absence of exposure to conspecific calls and because of the strong human social bonds formed, the animals were motivated to mimic human sounds that would not normally be considered a relevant model by these species. The age at which bonding occurred might also strongly influence whether a model is considered relevant (see also the section on timing below). In any case, these examples show us that the necessary peripheral morphology and the neural circuitry is present in seals, elephants and musk ducks, but that they will only mimic human speech sounds under extreme circumstances. As such, we must be careful to consider what an animal is 'capable' of doing if there is only an absence of evidence regarding the breadth of their vocal production learning capacity.

### (e) TIMING: when learning takes place

The age or seasonal context during which production learning can occur differs between species. In some birds, there is evidence that the formation of a template used for vocal learning may already start before they emerge from the egg. In the superb (*Malurus cyaneus*) and red-backed (*Malurus melanocephalus*) fairywrens, females give a specific call when incubating their eggs and chicks appear to incorporate these into their begging calls, suggesting that the calls are memorized *in ovo* [111–113]. For several songbirds, the phase for song learning is well defined. For example, zebra finches and white-crowned sparrows (*Zonotrichia leucophrys*) form their song memory (template) predominantly during a juvenile sensitive phase. They progressively match their later vocal output to this template and when the songs are crystallized and stereotyped the birds are then no longer able to learn new songs [114,115]. Other species, such as parrots are able to continue to learn new vocalization as adults, and some birds such as canaries or starlings go through a sensitive phase seasonally, learning new sounds or songs only during a specified annual window (e.g. [116,117]).

In mammals, the period during which vocal learning can take place has rarely been studied. Humans are able to learn novel sounds throughout their lifetime, but there appears to be a sensitive period for learning to distinguish or reproduce all aspects of the speech sounds of a language [118]. Humpback whales and bottlenose dolphins are clearly capable of

learning novel sounds throughout their lifetimes [119]. Single case reports, such as those for Hoover the seal [109] or Koshik the elephant [34], suggest that their production of learned vocalizations co-occurred with sexual maturation stages, but it appears likely in these cases that the learning of the auditory templates took place earlier in life as juveniles. Whether as mature adults they continued to add novel vocalizations is unknown. In a controlled experimental study, three grey seals less than a year of age were able to both learn and produce speech sounds while still juveniles [77]. For most non-human mammals, controlled experiments or long-term monitoring have not been performed and these are crucial to allow the determination of sensitive periods for vocal production learning.

### (f) TIMING: length of time learning takes

Modification of existing sounds can involve a rapid 'step-change' or gradual changes over time to produce the desired vocalization. Both of these can ultimately result in novel call types, but one happens immediately, the other, gradually. Memorizing a template and shaping vocal output accordingly may each take time, but a template may also be memorized rapidly in some species with initial vocal output already matching the model. For example, dolphins can be trained to imitate novel artificial sounds after the first exposure to a synthetic playback sound [40], and a nightingale (*Luscinia megarhynchos*) may reproduce a series of different songs it has heard only 20 times [120]. Converging on calls may also vary in the amount of exposure (trials) required. Some parrot species, such as the orange-fronted conures (*Aratinga canicularis*) can converge their contact calls towards a playback of a different contact call during a single playback session [87]. Conversely, in pale spear-nosed bats, many trials over approximately 30 days were required to learn to produce a small pitch shift [121].

### (g) TIMING: how long the learning is retained

In the previous section, we discussed whether animals rapidly learn to produce a new call type, or gradually learn to modify their call to approach the properties of the template. Related to this dimension is also the question of whether the internal auditory template that the animal is aiming to produce is stored in long term or only short-term memory. It is clear that in many classical cases of vocal production learning, such as songbird song learning, the internal auditory templates are maintained in long-term memory. For many songbirds, it has been shown that the learning process consists of two phases, a first phase of sensory learning, in which an auditory model is memorized as a template, and a second phase of sensory-motor learning in which the template guides the developing vocal outputs [90]. The songbird system thus puts strong emphasis on the long-term formation and use of a template.

Formation of a transient template in short-term memory would only allow for immediate matching of acoustic parameters or require repeated auditory input. If learning involves rapid production of new call types, such as in parrots and dolphins it is not always clear whether these templates are retained in long-term memory unless the resulting vocalizations are produced long term. Dolphins not only learn to imitate sounds immediately, but can be trained to use earlier mimicked sounds as object labels [40], indicating long-term storage of internal templates for these novel produced sounds. In the case of Alex, the African grey parrot,



it seems likely that the internal templates for the words he learnt were stored long-term as these sounds continued to be produced over a long period [74]. Tests of his comprehension and use of labels included objects that he might only see once per week [122], but he still could name them, so the internal templates were likely to have been retained at least this long. Differentiating between short- and long-term retention of learned templates will require experiments specifically designed to address this. For example, if matching to a sound that had been matched in the past is faster than the matching of a new sound, it would imply that some knowledge of the first template was retained in long-term memory. Exploring examples where short-term versus long-term retention of templates are observed could illuminate the mechanisms underlying these types of vocal production learning.

## 5. Grey areas in classification of vocal learning

Although we have endeavoured to produce a clear typology of vocal learning, some grey areas of classification remain. In some cases, it is not apparent if or when a learning process lies behind the observable behaviour. In other cases, the mechanisms behind behaviours identified as vocal variation, usage learning and/or production learning may be hard to distinguish between, or may even be co-occurring. In this section, we will discuss some of the outstanding 'grey areas' that hamper clear categorization of vocal learning across species.

In this paper, we have retained a reasonably strict definition that vocal production learning requires learning from an auditory model, and the formation of a template to which the vocal output is compared (see §4). However broader definitions of vocal production learning can be used in which the requirement for hearing an auditory model during the learning process is omitted [85]. An example of learning under this broader definition may be in Horsfield's bronze cuckoo (*Chalcites basalus*). The offspring of this brood parasitic species are reared by a foster species and after a few days produce begging calls that are quite similar to those of host young even though no host young are present. Experiments showed that this is the result of the gradual changing of a 'standard' begging call so that it becomes similar to host begging calls over a short period [123]. As there are no offspring of the foster species in the nest, there is no auditory model available to guide the change. The mechanism underlying the vocal change is suggested to be a type of operant learning [85]: parents are likely to provide more food when the chick begging call resembles what parents expect to hear from their offspring [123]. If the cuckoo receives limited responses to its begging, it starts to produce more variable calls, which might be considered improvisations on the cuckoo begging call. Those variants resulting in more food being delivered become more common and may be further varied so that it reaches a version which results in stable parental feeding. In this case, one might say that if there is a template, it is not in the mind of the learner, but in the mind of the rewarder [85]. Reward-driven changes in vocalizations have also been shown in systematic experiments on budgerigars [124], cowbirds [125], bats [33] and pinnipeds [126]. If the reward-driven changes are rather limited it is difficult to rule out that they were a consequence of usage learning, resulting in the production of a previously existing (but possibly rare) variant of a call. Conclusions on the production mechanism of

apparently novel or idiosyncratic sounds are especially hard to draw where comparisons to the natural repertoire and its inherent flexibility in wild conspecifics are lacking. If, however, reward-driven changes result in vocalizations well outside the known range of a species, such reward-driven vocal production learning may employ different mechanisms from those that rely on exposure to and subsequent matching of an auditory model by the learner.

Animals and humans may use a variety of learning mechanisms to learn to produce new sequences of vocalizations. In animal communication, learning sequences may involve simple linear arrangements of a small number of vocalizations (e.g. in primates [127–129]), or more complex arrangements such as songs that can involve combining calls or song syllables into repeating subunits of phrases or themes, some of which display complex syntax (e.g. birds [130–132], whales [133]). One example is Bengalese finch (*Lonchura striata* var. *domestica*) songs, which consist of syllable sequences that can be arranged in various orders, the sequence of which is affected by experience [134]. In some animal species, the rearrangement of vocalizations alters the functional response, which has been suggested to represent a simple combinatorial syntax (e.g. pied babblers (*Turdoides bicolor*) [135]). If an animal is reinforced to produce two vocalizations in a specific order, then associative learning can lead to sequence learning, and this would be categorized as usage learning. Sequence learning plays an important role in vocal production learning in some species as well. If learning a novel sequence of sounds leads to co-articulatory effects that modify the vocal output to produce a new sound [136], this could then be a case where both vocal usage and production learning could be argued. In humans and birds, there is evidence that the mechanism by which sequencing is learned is distinct from that which matches the spectral element structure to a template. In humans, learning the acoustic structure of phonemes appears to occur via a separate process to their sequencing in speech [137]. When birds learn a new song type, they have to acquire not only the elements that they need to produce but also the sequence and timing of these. There is also evidence that these two components are learned separately in songbirds, but that the processes are tightly integrated, similar to the processes that occur during human speech acquisition [134,138–140]. Currently, it is unclear whether this type of sequence learning uses the same neural structures as when the bird learns to give a call in a new social context or whether it shares components used in vocal production learning. Sequence learning is a key area for future study to understand processes by which it can take place and how they relate to, or overlap with the features and mechanisms underlying other aspects of vocal learning behaviour.

## 6. Mechanisms

The most accurate and useful classification system of vocal learning is one in which the behavioural categories are mirrored by mechanistic boundaries. Mechanistic understanding is thus key to resolving at least some of the grey areas in classification, such as those discussed above. For these reasons we discuss known and possible mechanisms underlying vocal learning, and in box 1 outline some key outstanding questions for future research that will lead to an improved typology and understanding of vocal learning.

**Box 1.** Key outstanding questions**Behaviour**

- During which life stages are animals capable of vocal learning and how (if at all) do the mechanisms that enable vocal production learning early in life differ from those that allow production learning later in life?
- How widespread is learning to use calls in a new sequence?
- Can the production of specific vocal rhythms be learned by animals?

**Mechanisms**

- What are the mechanistic underpinnings of the dimensions of vocal production learning?
- What are the mechanisms underlying the different phenomena described as usage learning and how do they relate to vocal production learning mechanisms?
- To what extent do usage learning and production learning share neural mechanisms?
- What are the mechanisms underlying sequence learning, and how do they relate to usage or production learning mechanisms?
- Is the process of modifying existing vocalizations mechanistically continuous or distinct from that of producing novel vocalizations?
- To what extent are auditory model guided and reward-based vocal modifications based on the same or different neural mechanisms?
- What neural and physiological constraints determine the vocal learning abilities of each species?
- How do neural control mechanisms of vocal production learning in laryngeal and non-laryngeal sound production compare?

**Evolution**

- How widespread are different types of vocal learning and how conserved over evolution are the mechanisms and circuits driving vocal learning in birds and mammals?
- What are the contexts in which (various types of) vocal learning may have arisen (e.g. socioecological factors)
- What are the selection processes (e.g. group member or individual recognition, kin selection, inter- and intra-sexual selection) that shaped vocal learning abilities in different species?

Studies in songbirds and humans demonstrate that vocal learning exploits specialized circuitry, but it is not yet known to what extent mechanisms underlying modification of existing vocalizations and learning novel sound types are overlapping or discrete in the brain, or how they evolved. Tyack posits that fine tuning of the acoustic features of existing calls in a repertoire represents a limited form of vocal learning, while the acquisition of a completely novel call type represents complex vocal learning, and that these are distinct behaviours involving different brain circuitry, such that complex vocal learning may have been achieved by supplementing subcortical circuits with new cortical connectivity [5,17]. Petkov & Jarvis [4] suggest that these extremes represent minima and maxima of a stepwise continuum and that existing circuits have been gradually built upon to increase behavioural complexity [4,141]. A major challenge for future research will be to elucidate the neural mechanisms underlying the different aspects of vocal modifications and vocal learning.

**(a) Vocal usage learning**

In this framework, usage learning covers a range of behaviours (associating existing sounds with novel contexts, learning new sequences of existing elements or altering the timing of sound production) that may employ different mechanisms. Furthermore, it is currently unclear how distinct or overlapping neural circuits involved in vocal usage learning are with those needed for vocal production learning. Vocal usage learning is widespread, while most usage

learners seem to lack the capacity for vocal production learning, suggesting specialized circuitry. However, this does not mean that there is no overlap in circuitry between the components of these behaviours. For example, in the song control system of zebra finches the premotor nucleus RA (nucleus robustus arcopallialis) and the projections that it receives from another telencephalic nucleus, HVC (proper name), are involved in the production and timing of non-learned calls in signal exchanges [63,142], showing that usage and production learning can share some neural structures. Thus, it may be that there is some shared recruitment of brain regions or circuits to subserve vocal usage and production learning, but that production learning requires the development of further specialized circuitry, particularly because usage learning can occur without the need for auditory input or matching of an auditory template. Instead, it can occur via 'learning from success'—i.e. learning that using a specific call is more likely to achieve the desired response [5,8,123,125,143]. This makes it likely that different pathways are involved in error-correction in vocal usage and vocal production learning.

**(b) Vocal production learning**

Vocal production learning requires exposure to an auditory model, formation of a memory of the target sound (template), control over anatomical structures to produce an approximation of the desired sound, and auditory feedback to perceive and adjust vocal output as needed. Neural circuits that control the perception, memorization, template matching

and motor production of learned vocalizations crucial for vocal learning are likely to have been built upon existing circuits over evolution. Although neural contributions to this complex trait must necessarily involve distributed circuitry throughout the brain, a few key pathways currently stand out as being specialized in vocal production learners [104,141].

Songbirds provide the best mechanistic understanding of the neural mechanisms involved in vocal production learning, in particular how model sounds are learned to form an auditory template and how auditory feedback during attempts to copy the sound allows the bird to converge on a model. In birds, the caudomedial nidopallium (NCM) has been proposed as a candidate site where tutor song memories are stored (reviewed in [144]). The NCM is an auditory region in the songbird ascending auditory pathway that, along with other forebrain and pallial nuclei, are viewed as analogous to the mammalian auditory cortex [145]. The NCM and another forebrain auditory region, the caudomedial mesopallium (CMM) are specialized for auditory learning and memory of conspecific vocalizations [146–150]. In an avian vocal mimic (budgerigar), vocalizations by conspecifics activate the NCM and CMM even when the sounds made by the conspecifics are unusual (e.g. a conspecific mimicking human speech [151]). In particular, data showing that activation of NCM predicts the accuracy with which zebra finch tutor songs are imitated [152–156] suggest that the precision of song imitation depends greatly upon the fidelity of memory encoding and consolidation of tutor song in this region. However, it has also been suggested that memories used to guide the vocal production of a learned auditory template are separately stored from those used to perceive and remember other conspecifics [157]. Instead of being stored as an auditory template in NCM, auditory templates that the bird is trying to produce are mapped on to and encoded in vocal-motor circuits (see [157–159]). In line with the idea that NCM is not involved in song imitation, lesions to NCM in juvenile zebra finches before or after tutor song exposure did not prevent them from learning to imitate tutor songs [160,161]. Juvenile zebra finches could also be optogenetically ‘tutored’ to sing experimentally specified song element durations by stimulating activity of premotor nucleus HVC via an upstream auditory region (Nif), i.e. independent of NCM. This effect occurs even if they have the opportunity to learn from a live adult male tutor [162]. Discovering whether tutor song is also stored in motor circuits of female songbirds that do not sing may provide further clarification on this matter.

Auditory feedback is a necessary element for shaping vocalizations in vocal production learning. Conceptually, auditory feedback of one’s own vocalizations can be compared to a memorized auditory template (comparator-type models) or by mapping auditory memories of tutor song onto vocal-motor commands so that as the bird sings, the motor connections that produce the desired auditory feedback are strengthened (inverse-type models, e.g. [163], see [158,164] for overviews of inverse models in songbirds). Nonetheless, extensive connectivity between motor and auditory areas is necessary for neuronal integration of auditory–vocal information. In humans, a dorsal auditory–parietal–motor cortical circuit has much stronger connectivity via the arcuate fasciculus compared to non-human primates [165]. Vocal learning birds also show novel connectivity

between auditory and vocal learning nuclei [141], as well as connectivity between song motor regions and ventral tegmental area dopaminergic neurons that provide reinforcement signals contingent on auditory feedback to guide learning [166–169]. Such auditory–vocal integration circuits may link the perception of a model sound to the programming of a novel motor output and/or matching of the produced sound to the model.

In vocal learning birds and humans, a direct forebrain-motor control pathway has been identified that is thought to enhance control over the syrinx (birds) or larynx (humans) and facilitate learned vocalizations (but see also comments on dolphins in the peripheral mechanisms section, below). The laryngeal motor cortex is a specialized region of the ventral primary motor cortex that connects to primary motor neurons in the brainstem that in turn innervate laryngeal muscles [170,171]. Species such as cats, which are thought not to be vocal production learners [172], innervate these muscles via indirect connections between the motor cortex and primary motor neurons [173]. By contrast, direct, strong mono-synaptic connections have been identified in oscine songbirds that learn their songs and in humans, and it has been hypothesized that these are required for vocal production learning [174]. Testing this hypothesis more broadly requires studying a range of species with varying levels of vocal production learning. These auditory and vocal-motor connections are poorly studied in most other vocal learning species—particularly in vocal learning mammals. This is in part owing to the inaccessibility of many mammalian vocal learners, but also the practical challenges that come with performing tracing or magnetic resonance imaging based studies in large and/or aquatic animals. Thus, determining the presence or absence of these connections in other more tractable vocal learning mammals such as bats or seals, coupled to behavioural studies of vocal learning capabilities, will be important to test these hypotheses and determine the necessary neural prerequisites for the range of vocal learning abilities found across species.

### (c) Peripheral mechanisms

The ability to vocally match a memorized model is not only constrained by the fidelity of memories or the learning circuitry, but also by the capabilities of the vocal production apparatus. The production of complex and varied vocalizations in birds and mammals employs sound source (e.g. larynx or syrinx) and sound filtering (e.g. orofacial control with lips, tongue, nasal passages, jaws and/or beaks) [104]. The structure of the larynx is highly conserved across mammals, so peripheral constraints of this vocal apparatus are expected to be less important for determining the breadth of vocal production learning within mammals than are neural constraints. For example, non-human primates have the physiological capacity to produce many human speech-like sounds [175]. Yet in practice, this does not occur even in the context of intensive training attempts, which is hypothesized to be owing to a lack of the appropriate learning circuitry [174]. However, the larynx is not always employed in mammalian vocal production learning. Dolphins are among the best mammalian production learners but do not use the larynx to produce sound. Sound originates in a newly evolved structure in the upper nasal passages called the dorsal bursae complex that includes two pairs of phonic

lips that can produce two sounds independently at the same time [176,177].

The avian equivalent of the larynx is the syrinx. Bird groups show more differences in the structure of their syrinx and the number of muscles that can modify its shape, than is observed for the mammalian larynx. So, it is conceivable that owing to this, different bird species have greater differences in the range of possible sounds they can produce. As an example: some species have two sound sources, one in each primary bronchus [178,179]. This latter structure allows some species to make two different types of sounds at the same time (e.g. brown-headed cowbird, *Molothrus ater ater*), which cannot be done with just one source [180,181]. It is, therefore, important to consider the peripheral morphological properties of a species to understand the potential constraints on vocal learning abilities.

## 7. Outlook

In this paper, we have aimed to build on the strong foundations of Janik & Slater [2,3], Petkov & Jarvis [4], Tyack [5] and others (e.g. [6–10]) to define a framework for understanding vocal learning across species. We contrasted non-learned vocal variation with vocal usage and vocal production learning. We divided vocal production learning into dimensions to display the complexity of this behaviour and allow a more fine-grained understanding of the components of vocal learning. This framework illustrates clear differences in behaviour, even between the most widely cited examples of vocal production learning species (figure 1). For example, while both a parrot and a zebra finch are capable of performing vocal production learning, zebra finches require an extended sensory-motor learning phase, while parrots quickly mimic a new auditory model. Furthermore, while zebra finches copy a narrow set of sounds during a limited developmental period, parrots can learn a much greater range of sounds, and continue to do so throughout life. Thus, while both are avian vocal learners, there are striking differences in their learning. Extending such a classification system to all vocal learning species would allow informed selection of species in which to best explore the mechanisms underlying a specific aspect of the set of behaviours that make up the dimensions of vocal learning. Similarly, extending this classification system would show which dimensions are most common or most rare, and which frequently co-occur, which may shed light on mechanistic or evolutionary relationships between dimensions.

A typology is only as good as the data on which it is based, and it is clear that vocal learning suffers from a severe absence of evidence in most species studied, as well as outstanding ‘grey areas’ where classifications cannot be clearly made, as is evident from the question marks (?) displayed in figure 1. Indeed, outside of songbirds (or even among them) we should consider the data to be largely incomplete. Thus, we acknowledge that the classifications outlined herein are based on the current available evidence. For this reason, figure 1 is not meant to be exhaustive, but is meant to represent a classification of a small selection of species based on currently available knowledge. Through much needed naturalistic and experimental observations, future studies are likely to reveal a wider range of species with vocal learning abilities than currently identified. But we hope that, as new evidence comes to light, our framework

will make it possible to align these species with what we know now. It would also be useful (although challenging) if future studies were able to reveal which species lack certain abilities. In the meantime, we must be clear about the difference between what an animal is *capable* of doing versus what we have been able to *observe* it doing. This is perfectly exemplified by the cases of Hoover the seal and most recently Ripper the Australian musk duck [109,110]. These single animal observations were so striking that they were able to indicate vocal production learning capacity for a species; however, they might easily have gone undiscovered without the right conditions fortuitously arising and being reported upon.

Our framework classifies the behavioural dimensions of vocal learning, in order to select species for the study of mechanisms of learning, but this categorization is by necessity an approximation. The ‘true’ definitions will match the underlying mechanisms, which may not align perfectly with the current behavioural distinctions. As such, determining the underlying neural mechanisms and demonstrating which behavioural distinctions are well supported by such mechanistic evidence is ultimately needed. Understanding mechanisms will help us resolve at least some of the grey areas of classification and determine the most important behavioural features on which to focus. For example, it might help resolve what is a ‘novel’ call, or whether a small change in vocalization employs the same process as learning to produce a very large change. It could illustrate if an animal learning a species-typical call type is performing the same task as one learning to mimic human speech, or if these are actually distinct processes. It is crucial that such work be performed comparatively in various avian and mammalian species to determine if the observed mechanisms are a universal hallmark of the behaviour, or a feature of the clade or species under study [7]. Exploring the underlying mechanisms will also reveal how the behaviours have evolved—e.g. via building on existing brain circuitry in a continuous fashion to improve skills as opposed to a dramatic step-change to introduce a new ability. Currently, mechanisms have not been sufficiently explored in enough species to make these distinctions or to generalize on how mechanisms align with behaviour. As such this is a clear area where future research is needed (see Box 1). We hope that the framework outlined here will aid in the long-term goal of producing of an integrated, mechanistic understanding of vocal learning across species and shed light on the evolution of vocal learning, including that in humans.

**Data accessibility.** This article has no additional data.

**Authors’ contributions.** All authors contributed to discussions shaping the text and provided portions of the text. S.C.V. prepared the final manuscript and all authors revised, edited and approved the publication of the final text.

**Competing interests.** We declare we have no competing interests.

**Funding.** S.C.V. was supported by a Max Planck Research Group (MPRG), a Human Frontiers Science Program (HFSP) Research grant (grant no. RGP0058/2016) and a UKRI Future Leaders Fellowship (grant no. MR/T021985/1). P.L.T. was supported by US Office of Naval Research (ONR) grant nos N00014-18-1-2062 and N00014-20-1-2709. B.P.K. was supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 751356. V.C.B. was supported by the DK Cognition and Communication by the Austrian Science Fund (FWF) grant no. W1262-B29.



**Acknowledgements.** We are indebted to the participants of the Lorentz workshop 'Unifying Vocal Learning' organized by Sonja Vernes, Vincent Janik, Constance Scharff and Tecumseh Fitch (2019). We thank the Lorentz Center and staff for the excellent support that made this workshop possible. We thank the co-organisers and the

participants for their inspiring contributions. We particularly thank Cedric Boeckx, Steffen Hage, Kurt Hammerschmidt, Pedro Tiago Martins, Andrea Ravignani, Angela Stoeger, Morgan Wirthlin, Ine Alvarez van Tussenbroek and Erich Jarvis for helpful discussions and input.

## References

- Darwin C. 1871 *The descent of man, and selection in relation to sex*. New York, NY: D. Appleton and Company.
- Janik VM, Slater PJB. 2000 The different roles of social learning in vocal communication. *Anim. Behav.* **60**, 1–11. (doi:10.1006/anbe.2000.1410)
- Janik VM, Slater PJB. 1997 Vocal learning in mammals. *Adv. Study Behav.* **26**, 59–99. (doi:10.1016/S0065-3454(08)60377-0)
- Petkov CI, Jarvis ED. 2012 Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* **4**, 12. (doi:10.3389/fnevo.2012.00012)
- Tyack PL. 2020 A taxonomy for vocal learning. *Phil. Trans. R. Soc. B* **375**, 20180406. (doi:10.1098/rstb.2018.0406)
- Beecher MD, Brenowitz EA. 2005 Functional aspects of song learning in songbirds. *Trends Ecol. Evol.* **20**, 143–149. (doi:10.1016/j.tree.2005.01.004)
- Lattenkamp EZ, Vernes SC. 2018 Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Curr. Opin. Behav. Sci.* **21**, 209–215. (doi:10.1016/j.cobeha.2018.04.007)
- Fischer J, Hammerschmidt K. 2020 Towards a new taxonomy of primate vocal production learning. *Phil. Trans. R. Soc. B* **375**, 20190045. (doi:10.1098/rstb.2019.0045)
- Martins PT, Boeckx C. 2020 Vocal learning: beyond the continuum. *PLoS Biol.* **18**, e3000672. (doi:10.1371/journal.pbio.3000672)
- Nowicki S, Searcy WA. 2014 The evolution of vocal learning. *Curr. Opin. Neurobiol.* **28**, 48–53. (doi:10.1016/j.conb.2014.06.007)
- Barron AB, Hebets EA, Cleland TA, Fitzpatrick CL, Hauber ME, Stevens JR. 2015 Embracing multiple definitions of learning. *Trends Neurosci.* **38**, 405–407. (doi:10.1016/j.tins.2015.04.008)
- Breed MD, Moore J. 2011 *Animal behaviour*. Cambridge, MA: Academic Press.
- Ey E, Hammerschmidt K, Zinner D, Fischer J. 2008 Influences of environmental factors on vocal communication in baboons. *Folia Primatol.* **79**, 327–327.
- Mikula P, Valcu M, Brumm H, Bulla M, Forstmeier W, Petruskova T, Kempenaers B, Albrecht T. 2021 A global analysis of song frequency in passerines provides no support for the acoustic adaptation hypothesis but suggests a role for sexual selection. *Ecol. Lett.* **24**, 477–486. (doi:10.1111/ele.13662)
- Brumm H. 2013 *Animal communication and noise*. Berlin, Germany: Springer.
- Brumm H, Slabbekoorn H. 2005 Acoustic communication in noise. *Adv. Study Behav.* **35**, 151–209. (doi:10.1016/S0065-3454(05)35004-2)
- Tyack PL. 2016 Vocal learning and auditory-vocal feedback. *Vertebr. Sound Prod. Acoust. Commun.* **53**, 261–295. (doi:10.1007/978-3-319-27721-9\_9)
- Ruch H, Zurcher Y, Burkart JM. 2018 The function and mechanism of vocal accommodation in humans and other primates. *Biol. Rev. Camb. Philos. Soc.* **93**, 996–1013. (doi:10.1111/brv.12382)
- Hotchkin C, Parks S. 2013 The Lombard effect and other noise-induced vocal modifications: insight from mammalian communication systems. *Biol. Rev. Camb. Philos. Soc.* **88**, 809–824. (doi:10.1111/brv.12026)
- Ulanovsky N, Fenton MB, Tsoar A, Korine C. 2004 Dynamics of jamming avoidance in echolocating bats. *Proc. R. Soc. Lond. B* **271**, 1467–1475. (doi:10.1098/rspb.2004.2750)
- Smotherman M, Zhang S, Metzner W. 2003 A neural basis for auditory feedback control of vocal pitch. *J. Neurosci.* **23**, 1464–1477. (doi:10.1523/JNEUROSCI.23-04-01464.2003)
- Metzner W, Zhang S, Smotherman M. 2002 Doppler-shift compensation behavior in horseshoe bats revisited: auditory feedback controls both a decrease and an increase in call frequency. *J. Exp. Biol.* **205**, 1607–1616. (doi:10.1242/jeb.205.11.1607)
- Ey E, Pfeufferle D, Fischer J. 2007 Do age- and sex-related variations reliably reflect body size in non-human primate vocalizations? A review. *Primates* **48**, 253–267. (doi:10.1007/s10329-006-0033-y)
- Briefer E, Vannoni E, McElligott AG. 2010 Quality prevails over identity in the sexually selected vocalizations of an ageing mammal. *BMC Biol.* **8**, 35. (doi:10.1186/1741-7007-8-35)
- Ehret G, Haack B. 1984 Motivation and arousal influence sound-induced maternal pup-retrieving behavior in lactating house mice. *Z. Tierpsychol.* **65**, 25–39. (doi:10.1111/j.1439-0310.1984.tb00370.x)
- Owren MJ, Rendall D. 2001 Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evol. Anthropol.* **10**, 58–71. (doi:10.1002/evan.1014)
- Schrader L, Todt D. 1998 Vocal quality is correlated with levels of stress hormones in domestic pigs. *Ethology* **104**, 859–876. (doi:10.1111/j.1439-0310.1998.tb00036.x)
- Takahashi DY, Fenley AR, Ghazanfar AA. 2016 Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Phil. Trans. R. Soc. B* **371**, 20150370. (doi:10.1098/rstb.2015.0370)
- Takahashi DY, Liao DA, Ghazanfar AA. 2017 Vocal learning via social reinforcement by infant marmoset monkeys. *Curr. Biol.* **27**, 1844–1852. (doi:10.1016/j.cub.2017.05.004)
- Liao DA, Zhang YSS, Cai LLX, Ghazanfar AA. 2018 Internal states and extrinsic factors both determine monkey vocal production. *Proc. Natl Acad. Sci. USA* **115**, 3978–3983. (doi:10.1073/pnas.1722426115)
- Semple S, McComb K. 2000 Perception of female reproductive state from vocal cues in a mammal species. *Proc. R. Soc. Lond. B* **267**, 707–712. (doi:10.1098/rspb.2000.1060)
- Adret P. 1993 Vocal learning induced with operant techniques: an overview. *Neth. J. Zool.* **43**, 125–142. (doi:10.1163/156854293x00250)
- Lattenkamp EZ, Vernes SC, Wiegrebe L. 2018 Volitional control of social vocalisations and vocal usage learning in bats. *J. Exp. Biol.* **221**, jeb.180729. (doi:10.1242/jeb.180729)
- Stoeger AS, Mietchen D, Oh S, de Silva S, Herbst CT, Kwon S, Fitch WT. 2012 An Asian elephant imitates human speech. *Curr. Biol.* **22**, 2144–2148. (doi:10.1016/j.cub.2012.09.022)
- Stansbury AL, de Freitas M, Wu GM, Janik VM. 2015 Can a gray seal (*Halichoerus grypus*) generalize call classes? *J. Comp. Psychol.* **129**, 412–420. (doi:10.1037/a0039756)
- Seyfarth R, Cheney D. 2018 Pragmatic flexibility in primate vocal production. *Curr. Opin. Behav. Sci.* **21**, 56–61. (doi:10.1016/j.cobeha.2018.02.005)
- Curio E, Ernst U, Vieth W. 1978 Cultural transmission of enemy recognition: one function of mobbing. *Science* **202**, 899–901. (doi:10.1126/science.202.4370.899)
- Brecht KF, Hage SR, Gavrilov N, Nieder A. 2019 Volitional control of vocalizations in corvid songbirds. *PLoS Biol.* **17**, e3000375. (doi:10.1371/journal.pbio.3000375)
- Hage SR, Nieder A. 2013 Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Commun.* **4**, 2409. (doi:10.1038/ncomms3409)
- Richards DG, Wolz JP, Herman LM. 1984 Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*. *J. Comp. Psychol.* **98**, 10–28. (doi:10.1037/0735-7036.98.1.10)
- Hauser MD. 1989 Ontogenetic changes in the comprehension and production of vervet monkey (*Cercopithecus aethiops*) vocalizations. *J. Comp. Psychol.* **103**, 149–158. (doi:10.1037/0735-7036.103.2.149)
- Munn CA. 1986 Birds that cry wolf. *Nature* **319**, 143–145. (doi:10.1038/319143a0)
- Flower TP, Gribble M, Ridley AR. 2014 Deception by flexible alarm mimicry in an African bird. *Science* **344**, 513–516. (doi:10.1126/science.1249723)
- Levinson SC. 2016 Turn-taking in human communication: origins and implications for

- language processing. *Trends Cogn. Sci.* **20**, 6–14. (doi:10.1016/j.tics.2015.10.010)
45. Wells KD. 1977 Social-behavior of anuran amphibians. *Anim. Behav.* **25**, 666–693. (doi:10.1016/0003-3472(77)90118-X)
46. Pika S, Wilkinson R, Kendrick KH, Vernes SC. 2018 Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B* **285**, 20180598. (doi:10.1098/rspb.2018.0598)
47. Rivera-Caceres KD, Quiros-Guerrero E, Araya-Salas M, Searcy WA. 2016 Neotropical wrens learn new duet rules as adults. *Proc. R. Soc. B* **283**, 20161819. (doi:10.1098/rspb.2016.1819)
48. Maples EG, Haraway MM, Hutto CW. 1989 Development of coordinated singing in a newly formed siamang pair (*Hylobates syndactylus*). *Zoo Biol.* **8**, 367–378. (doi:10.1002/zoo.1430080407)
49. Dingess KA, Slater PJB. 2006 Antiphonal four-part synchronized chorusing in a Neotropical wren. *Biol. Lett.* **2**, 1–4. (doi:10.1098/rsbl.2005.0373)
50. Rivera-Caceres KD, Quiros-Guerrero E, Araya-Salas M, Templeton CN, Searcy WA. 2018 Early development of vocal interaction rules in a duetting songbird. *R. Soc. Open Sci.* **5**, 171791. (doi:10.1098/rsos.171791)
51. Koda H, Lemasson A, Oyakawa C, Rizaldi, Pamungkas J, Masataka N. 2013 Possible role of mother-daughter vocal interactions on the development of species-specific song in gibbons. *PLoS ONE* **8**, e71432. (doi:10.1371/journal.pone.0071432)
52. Stivers T *et al.* 2009 Universals and cultural variation in turn-taking in conversation. *Proc. Natl Acad. Sci. USA* **106**, 10 587–10 592. (doi:10.1073/pnas.0903616106)
53. Chow CP, Mitchell JF, Miller CT. 2015 Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc. R. Soc. B* **282**, 20150069. (doi:10.1098/rspb.2015.0069)
54. Miller PJO, Shapiro AD, Tyack PL, Solow AR. 2004 Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*. *Anim. Behav.* **67**, 1099–1107. (doi:10.1016/j.anbehav.2003.06.017)
55. Carter GG, Skowronski MD, Faure PA, Fenton B. 2008 Antiphonal calling allows individual discrimination in white-winged vampire bats. *Anim. Behav.* **76**, 1343–1355. (doi:10.1016/j.anbehav.2008.04.023)
56. Soltis J, Leong K, Savage A. 2005 African elephant vocal communication I: antiphonal calling behaviour among affiliated females. *Anim. Behav.* **70**, 579–587. (doi:10.1016/j.anbehav.2004.11.015)
57. O'Connell-Rodwell CE, Wood JD, Wyman M, Redfield S, Puria S, Hart LA. 2012 Antiphonal vocal bouts associated with departures in free-ranging African elephant family groups (*Loxodonta africana*). *Bioacoustics* **21**, 215–224. (doi:10.1080/09524622.2012.686166)
58. Demartsev V, Strandburg-Peshkin A, Ruffner M, Manser M. 2018 Vocal turn-taking in meerkat group calling sessions. *Curr. Biol.* **28**, 3661. (doi:10.1016/j.cub.2018.09.065)
59. Yosida S, Kobayashi K, Ikebuchi M, Ozaki R, Okanoya K. 2007 Antiphonal vocalization of a subterranean rodent, the naked mole-rat (*Heterocephalus glaber*). *Ethology* **113**, 703–710. (doi:10.1111/j.1439-0310.2007.01371.x)
60. Hage SR, Jurgens U. 2006 Localization of a vocal pattern generator in the pontine brainstem of the squirrel monkey. *Eur. J. Neurosci.* **23**, 840–844. (doi:10.1111/j.1460-9568.2006.04595.x)
61. Patel AD, Iversen JR, Bregman MR, Schulz I. 2009 Studying synchronization to a musical beat in nonhuman animals. *Ann. N Y Acad. Sci.* **1169**, 459–469. (doi:10.1111/j.1749-6632.2009.04581.x)
62. 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. *J. Comp. Psychol.* **127**, 412–427. (doi:10.1037/a0032345)
63. Benichov JI, Benezra SE, Vallentin D, Globerson E, Long MA, Tchernichovski O. 2016 The forebrain song system mediates predictive call timing in female and male zebra finches. *Curr. Biol.* **26**, 309–318. (doi:10.1016/j.cub.2015.12.037)
64. van der Aa J, Honing H, ten Cate C. 2015 The perception of regularity in an isochronous stimulus in zebra finches (*Taeniopygia guttata*) and humans. *Behav. Processes* **115**, 37–45. (doi:10.1016/j.beproc.2015.02.018)
65. Watkins WA, Schevill WE. 1977 Sperm whale codas. *J. Acoust. Soc. Am.* **62**, 1485–1490. (doi:10.1121/1.381678)
66. Rendell LE, Whitehead H. 2003 Vocal clans in sperm whales (*Physeter macrocephalus*). *Proc. R. Soc. Lond. B* **270**, 225–231. (doi:10.1098/rspb.2002.2239)
67. Backus RH, Schevill WE. 1966 Physeter clicks. In *Whales, dolphins, and porpoises* (ed. KS Norris), pp. 510–527. Berkeley, CA: University of California Press.
68. Lampen J, Jones K, McAuley JD, Chang SE, Wade J. 2014 Arrhythmic song exposure increases ZENK expression in auditory cortical areas and nucleus taeniae of the adult zebra finch. *PLoS ONE* **9**, e108841. (doi:10.1371/journal.pone.0108841)
69. Mathevon N, Casey C, Reichmuth C, Charrier I. 2017 Northern elephant seals memorize the rhythm and timbre of their rivals' voices. *Curr. Biol.* **27**, 2352–2356. (doi:10.1016/j.cub.2017.06.035)
70. Shapiro AD, Slater PJB, Janik VM. 2004 Call usage learning in gray seals (*Halichoerus grypus*). *J. Comp. Psychol.* **118**, 447–454. (doi:10.1037/0735-7036.118.4.447)
71. Marler P, Peters S. 1988 The role of song phonology and syntax in vocal learning preferences in the song sparrow, *Melospiza melodia*. *Ethology* **77**, 125–149. (doi:10.1111/j.1439-0310.1988.tb00198.x)
72. Podos J, Nowicki S. 2004 Performance limits on birdsong. In *Nature's music: the science of bird song* (eds P Marler, H Slaabekoom), pp. 318–342. Amsterdam, The Netherlands: Elsevier.
73. Seyfarth RM, Cheney DL. 1997 Behavioral mechanisms underlying vocal communication in nonhuman primates. *Anim. Learn. Behav.* **25**, 249–267. (doi:10.3758/Bf03199083)
74. Pepperberg IM. 1999 *The Alex studies: cognitive and communicative abilities of grey parrots*, 434p. Cambridge, MA: Harvard University Press.
75. Pepperberg IM, Wilcox SE. 2000 Evidence for a form of mutual exclusivity during label acquisition by grey parrots (*Psittacus erithacus*)? *J. Comp. Psychol.* **114**, 219–231. (doi:10.1037/0735-7036.114.3.219)
76. Todt D. 1975 Spontaneous recombinations of vocal patterns in parrots. *Naturwissenschaften* **62**, 399–400. (doi:10.1007/Bf00625362)
77. Stansbury AL, Janik VM. 2019 Formant modification through vocal production learning in gray seals. *Curr. Biol.* **29**, 2244. (doi:10.1016/j.cub.2019.05.071)
78. Baptista LF, Morton ML. 1981 Interspecific song acquisition by a white-crowned sparrow. *Auk* **98**, 383–385.
79. Deregnacourt S, Poirier C, Van der Kant A, Van der Linden A, Gahr M. 2013 Comparisons of different methods to train a young zebra finch (*Taeniopygia guttata*) to learn a song. *J. Physiol. Paris* **107**, 210–218. (doi:10.1016/j.jphysparis.2012.08.003)
80. Roseberry S, Hirsh-Pasek K, Parish-Morris J, Golinkoff RM. 2009 Live action: can young children learn verbs from video? *Child Dev.* **80**, 1360–1375. (doi:10.1111/j.1467-8624.2009.01338.x)
81. Poole JH, Tyack PL, Stoeger-Horwath AS, Watwood S. 2005 Animal behaviour: elephants are capable of vocal learning. *Nature* **434**, 455–456. (doi:10.1038/434455a)
82. Zann RA. 1996 *The zebra finch: a synthesis of field and laboratory studies*, xvi, 335p. Oxford, NY: Oxford University Press.
83. Goller M, Shizuka D. 2018 Evolutionary origins of vocal mimicry in songbirds. *Evol. Lett.* **2**, 417–426. (doi:10.1002/evl3.62)
84. Johnson KE, Clark CJ. 2020 Ontogeny of vocal learning in a hummingbird. *Anim. Behav.* **167**, 139–150. (doi:10.1016/j.anbehav.2020.07.010)
85. ten Cate C. 2021 Re-evaluating vocal production learning in non-oscine birds. *Phil. Trans. R. Soc. B* **376**, 20200249. (doi:10.1098/rstb.2020.0249)
86. Esser KH. 1994 Audio-vocal learning in a non-human mammal: the lesser spear-nosed bat *Phyllostomus discolor*. *Neuroreport* **5**, 1718–1720. (doi:10.1097/00001756-199409080-00007)
87. Balsby TJ, Bradbury JW. 2009 Vocal matching by orange-fronted conures (*Aratinga canicularis*). *Behav. Processes* **82**, 133–139. (doi:10.1016/j.beproc.2009.05.005)
88. Johnson K. 2019 *Vocal learning in the Costa's hummingbird*. Riverside, CA: UC Riverside.
89. Fripp D, Owen C, Quintana-Rizzo E, Shapiro A, Buckstaff K, Jankowski K, Wells R, Tyack P. 2005 Bottlenose dolphin (*Tursiops truncatus*) calves appear to model their signature whistles on the signature whistles of community members. *Anim. Cogn.* **8**, 17–26. (doi:10.1007/s10071-004-0225-z)
90. Brainard MS, Doupe AJ. 2002 What songbirds teach us about learning. *Nature* **417**, 351–358. (doi:10.1038/417351a)
91. Sayigh LS, Tyack PL, Wells RS, Scott MD. 1990 Signature whistles of free-ranging bottlenose-dolphins *Tursiops truncatus*: stability and mother offspring comparisons. *Behav. Ecol. Sociobiol.* **26**, 247–260. (doi:10.1007/BF00178318)
92. Loth A. 2017 *Individual recognition in bottlenose dolphins (Tursiops truncatus): from acoustic*

- distinctiveness in signature whistles to visual self-recognition*. St Andrews, UK: University of St Andrews.
93. Walloe S, Thomsen H, Balsby TJ, Dabelsteen T. 2015 Differences in short-term vocal learning in parrots, a comparative study. *Behaviour* **152**, 1433. (doi:10.1163/1568539x-00003286)
  94. Soha JA, Lohr B, Gill DE. 2009 Song development in the grasshopper sparrow, *Ammodramus savannarum*. *Anim. Behav.* **77**, 1479–1489. (doi:10.1016/j.anbehav.2009.02.021)
  95. Tyack P. 1986 Whistle repertoires of 2 bottle-nosed dolphins, *Tursiops truncatus*: mimicry of signature whistles. *Behav. Ecol. Sociobiol.* **18**, 251–257. (doi:10.1007/Bf00300001)
  96. King SL, Janik VM. 2013 Bottlenose dolphins can use learned vocal labels to address each other. *Proc. Natl Acad. Sci. USA* **110**, 13 216–13 221. (doi:10.1073/pnas.1304459110)
  97. King SL, Sayigh LS, Wells RS, Fellner W, Janik VM. 2013 Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proc. R. Soc. B* **280**, 20130053. (doi:10.1098/rspb.2013.0053)
  98. Lachlan RF, Anderson RC, Peters S, Searcy WA, Nowicki S. 2014 Typical versions of learned swamp sparrow song types are more effective signals than are less typical versions. *Proc. R. Soc. B* **281**, 20140252. (doi:10.1098/rspb.2014.0252)
  99. Spencer KA, Buchanan KL, Goldsmith AR, Catchpole CK. 2003 Song as an honest signal of developmental stress in the zebra finch (*Taeniopygia guttata*). *Horm. Behav.* **44**, 132–139. (doi:10.1016/S0018-506x(03)00124-7)
  100. Holveck MJ, de Castro ACV, Lachlan RF, ten Cate C, Riebel K. 2008 Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. *Behav. Ecol.* **19**, 1267–1281. (doi:10.1093/beheco/arn078)
  101. Sewall KB, Anderson RC, Soha JA, Peters S, Nowicki S. 2018 Early life conditions that impact song learning in male zebra finches also impact neural and behavioral responses to song in females. *Dev. Neurobiol.* **78**, 785–798. (doi:10.1002/dneu.22600)
  102. Boogert NJ, Lachlan RF, Spencer KA, Templeton CN, Farine DR. 2018 Stress hormones, social associations and song learning in zebra finches. *Phil. Trans. R. Soc. B* **373**, 20170290. (doi:10.1098/rstb.2017.0290)
  103. Kriengwatana BP. 2019 Learning strategies and the social brain: missing elements in the link between developmental stress, song and cognition? *Integr. Zool.* **14**, 158–171. (doi:10.1111/1749-4877.12379)
  104. Fitch WT. 2000 The evolution of speech: a comparative review. *Trends Cogn. Sci.* **4**, 258–267. (doi:10.1016/S1364-6613(00)01494-7)
  105. Green JR, Moore CA, Higashikawa M, Steeve RW. 2000 The physiologic development of speech motor control: lip and jaw coordination. *J. Speech Lang. Hear Res.* **43**, 239–255. (doi:10.1044/jslhr.4301.239)
  106. Dalziell AH, Welbergen JA, Igic B, Magrath RD. 2015 Avian vocal mimicry: a unified conceptual framework. *Biol. Rev. Camb. Philos. Soc.* **90**, 643–668. (doi:10.1111/brv.12129)
  107. Zann R, Dunstan E. 2008 Mimetic song in superb lyrebirds: species mimicked and mimetic accuracy in different populations and age classes. *Anim. Behav.* **76**, 1043–1054. (doi:10.1016/j.anbehav.2008.05.021)
  108. Dalziell AH, Welbergen JA. 2016 Elaborate mimetic vocal displays by female superb lyrebirds. *Front. Ecol. Evol.* **4**, 34. (doi:10.3389/fevo.2016.00034)
  109. Ralls K, Fiorelli P, Gish S. 1985 Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. *Can. J. Zool.* **63**, 1050–1056. (doi:10.1139/z85-157)
  110. ten Cate C, Fullagar P. 2021 Vocal imitations and production learning by Australian musk ducks (*Biziura lobata*). *Phil. Trans. R. Soc. B* **376**, 20200243. (doi:10.1098/rstb.2020.0243)
  111. Colombelli-Negrel D, Hauber ME, Robertson J, Salloway FJ, Hoi H, Griggio M, Kleindorfer S. 2012 Embryonic learning of vocal passwords in superb fairy-wrens reveals intruder cuckoo nestlings. *Curr. Biol.* **22**, 2155–2160. (doi:10.1016/j.cub.2012.09.025)
  112. Colombelli-Negrel D, Kleindorfer S. 2017 Prenatal environment affects embryonic response to song. *Biol. Lett.* **13**, 20170302. (doi:10.1098/rsbl.2017.0302)
  113. Dowling JL, Colombelli-Negrel D, Webster MS. 2016 Kin signatures learned in the egg? Red-backed fairy-wren songs are similar to their mother's in-nest calls and songs. *Front. Ecol. Evol.* **4**, 48. (doi:10.3389/fevo.2016.00048)
  114. Wilbrecht L, Nottebohm F. 2003 Vocal learning in birds and humans. *Ment. Retard. Dev. Disabil. Res. Rev.* **9**, 135–148. (doi:10.1002/mrdd.10073)
  115. Marler P. 1997 Three models of song learning: evidence from behavior. *J. Neurobiol.* **33**, 501–516. (doi:10.1002/(SICI)1097-4695(19971105)33:5<501::AID-NEU2>3.0.CO;2-8)
  116. Nottebohm F, Nottebohm ME, Crane LA, Wingfield JC. 1987 Seasonal changes in gonadal hormone levels of adult male canaries and their relation to song. *Behav. Neural. Biol.* **47**, 197–211. (doi:10.1016/s0163-1047(87)90327-x)
  117. Voigt C, Leitner S. 2008 Seasonality in song behaviour revisited: seasonal and annual variants and invariants in the song of the domesticated canary (*Serinus canaria*). *Horm. Behav.* **54**, 373–378. (doi:10.1016/j.yhbeh.2008.05.001)
  118. Oyama S. 1976 Sensitive period for acquisition of a non-native phonological system. *J. Psycholinguist. Res.* **5**, 261–283. (doi:10.1007/Bf01067377)
  119. Janik VM. 2014 Cetacean vocal learning and communication. *Curr. Opin. Neurobiol.* **28C**, 60–65. (doi:10.1016/j.conb.2014.06.010)
  120. Hultsch H, Todt D. 1989 Memorization and reproduction of songs in nightingales (*Luscinia megarhynchos*): evidence for package formation. *J. Comp. Physiol. A-Sens. Neural. Behav. Physiol.* **165**, 197–203. (doi:10.1007/Bf00619194)
  121. Lattenkamp EZ, Vernes SC, Wiegrebe L. 2020 Vocal production learning in the pale spear-nosed bat, *Phyllostomus discolor*. *Biol. Lett.* **16**, 20190928. (doi:10.1098/rsbl.2019.0928)
  122. Pepperberg IM. 1990 Cognition in an African gray parrot (*Psittacus erithacus*): further evidence for comprehension of categories and labels. *J. Comp. Psychol.* **104**, 41–52. (doi:10.1037/0735-7036.104.1.41)
  123. Langmore NE, Maurer G, Adcock GJ, Kilner RM. 2008 Socially acquired host-specific mimicry and the evolution of host races in Horsfield's bronze-cuckoo *Chalcites basalus*. *Evolution* **62**, 1689–1699. (doi:10.1111/j.1558-5646.2008.00405.x)
  124. Manabe K, Dooling RJ. 1997 Control of vocal production in budgerigars (*Melopsittacus undulatus*): selective reinforcement, call differentiation, and stimulus control. *Behav. Processes* **41**, 117–132. (doi:10.1016/S0376-6357(97)00041-7)
  125. West MJ, King AP. 1988 Female visual displays affect the development of male song in the cowbird. *Nature* **334**, 244–246. (doi:10.1038/334244a0)
  126. Schusterman RJ, Reichmuth C. 2008 Novel sound production through contingency learning in the Pacific walrus (*Odobenus rosmarus divergens*). *Anim. Cogn.* **11**, 319–327. (doi:10.1007/s10071-007-0120-5)
  127. Ouattara K, Lemasson A, Zuberbuhler K. 2009 Campbell's monkeys use affixation to alter call meaning. *PLoS ONE* **4**, e7808. (doi:10.1371/journal.pone.0007808)
  128. Coye C, Ouattara K, Zuberbuhler K, Lemasson A. 2015 Suffixation influences receivers' behaviour in non-human primates. *Proc. R. Soc. B* **282**, 20150265. (doi:10.1098/rspb.2015.0265)
  129. Marshall AJ, Wrangham RW, Arcadi AC. 1999 Does learning affect the structure of vocalizations in chimpanzees? *Anim. Behav.* **58**, 825–830. (doi:10.1006/anbe.1999.1219)
  130. Catchpole C, Slater PJB. 2008 *Bird song: biological themes and variations*, 2nd edn. Cambridge, UK: Cambridge University Press.
  131. Suzuki TN, Wheatcroft D, Griesser M. 2016 Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* **7**, 10986. (doi:10.1038/ncomms10986)
  132. Suzuki TN, Wheatcroft D, Griesser M. 2017 Wild birds use an ordering rule to decode novel call sequences. *Curr. Biol.* **27**, 2331. (doi:10.1016/j.cub.2017.06.031)
  133. Payne RS, McVay S. 1971 Songs of humpback whales. *Science* **173**, 585–597. (doi:10.1126/science.173.3997.585)
  134. James LS, Sun H, Wada K, Sakata JT. 2020 Statistical learning for vocal sequence acquisition in a songbird. *Sci. Rep.* **10**, 2248. (doi:10.1038/s41598-020-58983-8)
  135. Engesser S, Ridley AR, Townsend SW. 2016 Meaningful call combinations and compositional processing in the southern pied babbler. *Proc. Natl Acad. Sci. USA* **113**, 5976–5981. (doi:10.1073/pnas.1600970113)
  136. Recasens D. 2018 *Coarticulation*. The Oxford research encyclopedia of Linguistics. Oxford, UK: Oxford University Press.
  137. Guenther FH. 2016 *Neural control of speech*. Cambridge, MA: The MIT Press.



138. Lipkind D, Geambasu A, Levelt CC. 2020 The development of structured vocalizations in songbirds and humans: a comparative analysis. *Top. Cogn. Sci.* **12**, 894–909. (doi:10.1111/tops.12414)
139. Tchernichovski O, Lipkind D. 2017 Animal communication: origins of sequential structure in birdsong. *Curr. Biol.* **27**, R1268–R1269. (doi:10.1016/j.cub.2017.10.041)
140. James LS, Sakata JT. 2017 Learning biases underlie ‘universals’ in avian vocal sequencing. *Curr. Biol.* **27**, 3676–3682. (doi:10.1016/j.cub.2017.10.019)
141. Jarvis ED. 2019 Evolution of vocal learning and spoken language. *Science* **366**, 50–54. (doi:10.1126/science.aax0287)
142. Ter Maat A, Trost L, Sagunsky H, Seltmann S, Gahr M. 2014 Zebra finch mates use their forebrain song system in unlearned call communication. *PLoS ONE* **9**, e109334. (doi:10.1371/journal.pone.0109334)
143. Marler P, Nelson DA. 1993 Action-based learning: a new form of developmental plasticity in bird song. *Neth. J. Zool.* **43**, 91–103. (doi:10.1163/156854293X00232)
144. Bolhuis JJ, Moorman S. 2015 Birdsong memory and the brain: in search of the template. *Neurosci. Biobehav. Rev.* **50**, 41–55. (doi:10.1016/j.neubiorev.2014.11.019)
145. Pinaud R, Terleph TA. 2008 A songbird forebrain area potentially involved in auditory discrimination and memory formation. *J. Biosci.* **33**, 145–155. (doi:10.1007/s12038-008-0030-y)
146. Mello C, Nottebohm F, Clayton D. 1995 Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene’s response to that song in zebra finch telencephalon. *J. Neurosci.* **15**, 6919–6925. (doi:10.1523/JNEUROSCI.15-10-06919.1995)
147. MacDougall-Shackleton SA, Hulse SH, Ball GF. 1998 Neural correlates of singing behavior in male zebra finches (*Taeniopygia guttata*). *J. Neurobiol.* **36**, 421–430. (doi:10.1002/(SICI)1097-4695(19980905)36:3<421::AID-NEU9>3.0.CO;2-Y)
148. Gentner TQ, Hulse SH, Duffy D, Ball GF. 2001 Response biases in auditory forebrain regions of female songbirds following exposure to sexually relevant variation in male song. *J. Neurobiol.* **46**, 48–58. (doi:10.1002/1097-4695(200101)46:1<48::AID-NEU5>3.0.CO;2-3)
149. Chew SJ, Vicario DS, Nottebohm F. 1996 A large-capacity memory system that recognizes the calls and songs of individual birds. *Proc. Natl Acad. Sci. USA* **93**, 1950–1955. (doi:10.1073/pnas.93.5.1950)
150. Mello CV, Vicario DS, Clayton DF. 1992 Song presentation induces gene-expression in the songbird forebrain. *Proc. Natl Acad. Sci. USA* **89**, 6818–6822. (doi:10.1073/pnas.89.15.6818)
151. Eda-Fujiwara H *et al.* 2012 Localized brain activation related to the strength of auditory learning in a parrot. *PLoS ONE* **7**, e38803. (doi:10.1371/journal.pone.0038803)
152. Bolhuis JJ, Zijlstra GGO, den Boer-Visser AM, Van der Zee EA. 2000 Localized neuronal activation in the zebra finch brain is related to the strength of song learning. *Proc. Natl Acad. Sci. USA* **97**, 2282–2285. (doi:10.1073/pnas.030539097)
153. Gobes SM, Zandbergen MA, Bolhuis JJ. 2010 Memory in the making: localized brain activation related to song learning in young songbirds. *Proc. R. Soc. B* **277**, 3343–3351. (doi:10.1098/rspb.2010.0870)
154. Terpstra NJ, Bolhuis JJ, Riebel K, van der Burg JMM, den Boer-Visser AM. 2006 Localized brain activation specific to auditory memory in a female songbird. *J. Comp. Neurol.* **494**, 784–791. (doi:10.1002/cne.20831)
155. Phan ML, Pytte CL, Vicario DS. 2006 Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proc. Natl Acad. Sci. USA* **103**, 1088–1093. (doi:10.1073/pnas.0510136103)
156. London SE, Clayton DF. 2008 Functional identification of sensory mechanisms required for developmental song learning. *Nat. Neurosci.* **11**, 579–586. (doi:10.1038/nm.2103)
157. Ikeda MZ, Trusel M, Roberts TF. 2020 Memory circuits for vocal imitation. *Curr. Opin. Neurobiol.* **60**, 37–46. (doi:10.1016/j.conb.2019.11.002)
158. Lewandowski B, Vysotski A, Hahnloser RH, Schmidt M. 2013 At the interface of the auditory and vocal motor systems: Nlf and its role in vocal processing, production and learning. *J. Physiol. Paris* **107**, 178–192. (doi:10.1016/j.jphysparis.2013.04.001)
159. Bertram R, Hyson RL, Brunick AJ, Flores D, Johnson F. 2020 Network dynamics underlie learning and performance of birdsong. *Curr. Opin. Neurobiol.* **64**, 119–126. (doi:10.1016/j.conb.2020.04.004)
160. Canopoli A, Zai AT, Hahnloser RH. 2016 Lesions of a higher auditory brain area during a sensorimotor period do not impair birdsong learning. *Matters* **2**, e201603000018. (doi:10.19185/matters.201603000018)
161. Canopoli A, Zai AT, Hahnloser RH. 2017 Bilateral neurotoxic lesions in NCM before tutoring onset do not prevent successful tutor song learning. *Matters* **3**, e201612000007. (doi:10.19185/matters.201612000007)
162. Zhao WC, Garcia-Oscos F, Dinh D, Roberts TF. 2019 Inception of memories that guide vocal learning in the songbird. *Science* **366**, 83. (doi:10.1126/science.aaw4226)
163. Giret N, Kornfeld J, Ganguli S, Hahnloser RH. R. 2014 Evidence for a causal inverse model in an avian cortico-basal ganglia circuit. *Proc. Natl Acad. Sci. USA* **111**, 6063–6068. (doi:10.1073/pnas.1317087111)
164. Burgess JD, Lum JA. G., Hohwy J, Enticott PG. 2017 Echoes on the motor network: how internal motor control structures afford sensory experience. *Brain Struct. Funct.* **222**, 3865–3888. (doi:10.1007/s00429-017-1484-1)
165. Balezau F, Wilson B, Gallardo G, Dick F, Hopkins W, Anwander A, Friederici AD, Griffiths TD, Petkov CI. 2020 Primate auditory prototype in the evolution of the arcuate fasciculus. *Nat. Neurosci.* **23**, 611–614. (doi:10.1038/s41593-020-0623-9)
166. Andalman AS, Fee MS. 2009 A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc. Natl Acad. Sci. USA* **106**, 12 518–12 523. (doi:10.1073/pnas.0903214106)
167. Gadagkar V, Puzerey PA, Chen R, Baird-Daniel E, Farhang AR, Goldberg JH. 2016 Dopamine neurons encode performance error in singing birds. *Science* **354**, 1278–1282. (doi:10.1126/science.aah6837)
168. Chen RD, Puzerey PA, Roeser AC, Riccelli TE, Podury A, Maher K, Farhang AR, Goldberg JH. 2019 Songbird ventral pallidum sends diverse performance error signals to dopaminergic midbrain. *Neuron* **103**, 266. (doi:10.1016/j.neuron.2019.04.038)
169. Kearney MG, Warren TL, Hisey E, Qi JX, Mooney R. 2019 Discrete evaluative and premotor circuits enable vocal learning in songbirds. *Neuron* **104**, 559. (doi:10.1016/j.neuron.2019.07.025)
170. Simonyan K. 2014 The laryngeal motor cortex: its organization and connectivity. *Curr. Opin. Neurobiol.* **28**, 15–21. (doi:10.1016/j.conb.2014.05.006)
171. Simonyan K, Horwitz B. 2011 Laryngeal motor cortex and control of speech in humans. *Neuroscientist* **17**, 197–208. (doi:10.1177/1073858410386727)
172. Romand R, Ehret G. 1984 Development of sound production in normal, isolated, and deafened kittens during the first postnatal months. *Dev. Psychobiol.* **17**, 629–649. (doi:10.1002/dev.420170606)
173. Kuypers HG. 1958 An anatomical analysis of corticobulbar connexions to the pons and lower brain stem in the cat. *J. Anat.* **92**, 198–218.
174. Fitch WT. 2018 The biology and evolution of speech: a comparative analysis. *Annu. Rev. Linguist.* **4**, 255–279. (doi:10.1146/annurev-linguistics-011817-045748)
175. Fitch WT, de Boer B, Mathur N, Ghazanfar AA. 2016 Monkey vocal tracts are speech-ready. *Sci. Adv.* **2**, e1600723. (doi:10.1126/sciadv.1600723)
176. Cranford TW. 2011 Biosonar sources in odontocetes: considering structure and function. *J. Exp. Biol.* **214**, 1403–1404. (doi:10.1242/jeb.053660)
177. Cozzi B, Huguenberger S, Oelschläger H. 2017 Chapter 1. In *Anatomy of dolphins: insights into body structure and function*, pp. 1–20. Cambridge, MA: Academic Press.
178. Garcia SM, Kopuchian C, Fuxjager MJ, Riede T, Goller F. 2017 Evolution of diverse song: functional morphology of the avian syrinx and motor control in suboscines and oscines. *Integr. Comp. Biol.* **57**, E268–E268.
179. Suthers RA. 1992 Lateralization of sound production and motor action on the left and right sides of the syrinx during bird song. In *14th Int. Congress on Acoustics, Proc., 3–10 September 1992, Beijing, China*, vols. 1–4, pp. 1333–1334.
180. Allan SE, Suthers RA. 1994 Lateralization and motor stereotypy of song production in the brown-headed cowbird. *J. Neurobiol.* **25**, 1154–1166. (doi:10.1002/neu.480250910)
181. Zollinger SA, Suthers RA. 2004 Motor mechanisms of a vocal mimic: implications for birdsong production. *Proc. R. Soc. Lond. B* **271**, 483–491. (doi:10.1098/rspb.2003.2598)