



Opinion piece

Cite this article: Lattenkamp EZ, Hörpel SG, Mengede J, Firzlaff U. 2021 A researcher's guide to the comparative assessment of vocal production learning. *Phil. Trans. R. Soc. B* **376**: 20200237.

<https://doi.org/10.1098/rstb.2020.0237>

Accepted: 8 March 2021

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

Subject Areas:

neuroscience, behaviour

Keywords:

cross-species comparison, vocal production learning, comparative VPL assessment, vocal learning quality

Author for correspondence:

Ella Z. Lattenkamp

e-mail: ella.lattenkamp@evobio.eu

A researcher's guide to the comparative assessment of vocal production learning

Ella Z. Lattenkamp^{1,2}, Stephen G. Hörpel^{2,3}, Janine Mengede² and Uwe Firzlaff³

¹Division of Neurobiology, Department of Biology II, LMU Munich, Germany

²Neurogenetics of Vocal Communication Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

³Department of Animal Sciences, Chair of Zoology, TU Munich, Germany

EZL, 0000-0002-8928-8770; SGH, 0000-0001-6104-8970; UF, 0000-0003-2803-3587

Vocal production learning (VPL) is the capacity to learn to produce new vocalizations, which is a rare ability in the animal kingdom and thus far has only been identified in a handful of mammalian taxa and three groups of birds. Over the last few decades, approaches to the demonstration of VPL have varied among taxa, sound production systems and functions. These discrepancies strongly impede direct comparisons between studies. In the light of the growing number of experimental studies reporting VPL, the need for comparability is becoming more and more pressing. The comparative evaluation of VPL across studies would be facilitated by unified and generalized reporting standards, which would allow a better positioning of species on any proposed VPL continuum. In this paper, we specifically highlight five factors influencing the comparability of VPL assessments: (i) comparison to an acoustic baseline, (ii) comprehensive reporting of acoustic parameters, (iii) extended reporting of training conditions and durations, (iv) investigating VPL function via behavioural, perception-based experiments and (v) validation of findings on a neuronal level. These guidelines emphasize the importance of comparability between studies in order to unify the field of vocal learning.

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

1. The need for comparability and unity across reports of vocal production learning

Vocal production learning (VPL) has been described as 'instances where [acoustic] signals themselves are modified in form as a result of experience with those of other individuals' [1]. The definition of this complex behavioural trait, which is thought to be one of the evolutionary prerequisites for the human capacity for speech, has been changed repeatedly and redefined in a number of studies. Among others, VPL has been described as 'matching', 'imitating', 'copying', 'reproducing', 'resembling' and 'vocally mimicking' conspecific, heterospecific or artificially generated acoustic signals. The fickle and varied nature of these descriptions is based in part on the diversity of its expression and, additionally, on the heterogeneity of its measurements. As the delimitation of VPL is in a phase of redefinition, as evidenced by this special issue, the evidence we are looking at in order to inform these definitions should at least be comparable, reliable and uniform. However, in different experimental studies describing VPL, the same terminology is often used to describe sometimes drastically different findings. In order to compare findings of studies investigating multidimensional behavioural traits such as VPL, not only within and between species, but also independent of the definition used at the time of the study, it is of utmost importance to make the reporting of the findings as comparable and explicit as possible. In this paper, we highlight the importance of such comparability between studies and provide guidelines for the unified reporting of VPL

evidence. The implementation of these guidelines will facilitate the organization of vocal learners within the vocal learning parameter space and allow the comparative assessment of VPL to adapt flexibly according to the definition used.

We want to give two examples of the difficulty of the comparative assessment of VPL. The first example illustrates the discrepancies in the demonstration of VPL via the imitation of human speech. Examples for the imitation of human speech include studies in elephants [2], seals [3], songbirds [4], parrots [5] and cetaceans (belugas [6,7] and killer whales [8]). Some of these studies include the results of months of extensive training, while others report spontaneous mimicry. Assessments of the similarity between animal vocalization and human-produced acoustic targets vary strongly among these studies. A common evaluation strategy is to enlist human raters to either transcribe the recordings [2] or judge acoustic similarity between the recordings and the target sound [6,7]. Furthermore, it is typical to assess the similarity between tutor and tutee vocalizations based solely on visual inspection of spectrograms [5,7]. However, repeatable acoustic parameter extraction and comparison, such as discriminant functional analyses or distance matrices based on specific extracted parameters [2,3], are often lacking.

The second example illustrates the difficulty of comparing VPL studies within animal clades. The capacity for VPL in bats has thus far been indicated for a handful of species [9]. The evidence for this capacity, however, is again provided by a variety of different study designs and reported parameters. While some studies used experimental designs to specifically modify social group structure (such as isolation studies [10] or transfer studies [11]), other studies focused on vocal adjustment in response to a playback [12–15] or on recordings in the wild [16–18]. These studies also vary in the main parameters investigated to assess VPL in bats. While some studies focus on the fundamental frequency [10,15,16], others focus on bandwidth [11,12], or spectral centroid frequency [14], or used discriminant function analyses to assess a number of parameters in combination [18].

These two approaches to the demonstration of VPL (i.e. human speech mimicry in a variety of taxa and studying different expressions of VPL in one taxon) have been conducted with considerable variation in both study design and reported parameters. While all of these studies claim the demonstration of VPL, the presented evidence has varied in its success at convincing the scientific community. This scepticism is rooted in the inability to compare the evidence against one another. As applying different approaches to the demonstration of VPL is, of course, crucial for such a diverse field of study, the lack of comparability is a key obstacle in the field of VPL research. To allow the inter- and intraspecific comparison of VPL capacity in the future, measured and reported parameters need to be comparable across a wide array of studies.

The assessment of VPL capacity can often be reduced to a test of similarity between tutee and tutor (conspecific, heterospecific or artificial) and in the long run also to the judgement of qualities such as novelty and complexity of the observed vocal imitations. Our assessment of either vocal imitation of single acoustic parameters or the sum of all parameters is often concerned with the question of whether the individual or species has the capacity to learn (to imitate) an acoustic signal precisely (VPL quality). But what do VPL studies mean and report when the precision or quality of imitation

is described? In experimental studies, the VPL trait is often evaluated on the basis of comparisons within an acoustic parameter space. Especially for zebra finches, the currently most studied VPL model species [19], several different automatic algorithms have been developed to assess the similarity of their calls or songs [20,21]. However, these are often quite specific for their focal species and dependent on laboratory recording conditions. In the wild, VPL is often more subtle and harder to demonstrate due to the lack of controlled recording conditions. Most importantly, in the wild, not only the change within the acoustic parameter space is essential, but also the behavioural response to and the social reinforcement of the trait are important for the comprehensive assessment of VPL. Therefore, we focus here on two types of demonstration of VPL: the bioacoustic, analytical evaluation and the behavioural or neuronal evaluation of acoustic differences/similarities. The evaluation of VPL can, thus, be twofold and concern either the acoustic parameter space and/or the behavioural decision-making/perceptual space. Both spaces can be modified due to VPL and both can be assessed when VPL quality is studied. In the following, we present guidelines for reported acoustic parameters and additional studies, which help to comprehensively describe a species' capacity for VPL, assess possibilities for external validation, and ultimately make the findings available for cross-species comparisons of VPL.

2. The acoustic parameter space

(a) The need for a robust baseline in order to assess vocal production learning precision (and novelty)

For the assessment of VPL, a species' typical acoustic variation needs to be considered as baseline. Only by considering the species-specific vocal repertoire and the inter- and intra-individual vocal variation can we discriminate learned and experience-independent vocalizations and, moreover, assess the placement of a 'new' vocalization within the species' distribution of the acoustic feature in focus. Judging novelty is one of the hardest tasks in the field of vocal learning; nevertheless, it is often considered one of the clearest distinctions between vocal usage and VPL [1,22]. But when is something novel and how much does the novel signal need to vary from existing calls? These questions need to be answered considering the species-specific vocal repertoire. The variation of parameters in the acoustic environment and a subject's pre-exposure repertoire can give us an idea of the variability and the importance of different acoustic parameters (i.e. of behavioural relevance for the species). For example, if a 'novel vocalization' is defined as existing outside of the acoustic 'feature space' of the natural variety of the species, the species/population mean needs to be consulted as a ground truth. Knowledge about a species' typical vocal variation also has implications for experimental study design. The generation of artificial acoustic targets for imitation studies needs to be informed by the species' vocal baseline to avoid either overlapping with the pre-existing repertoire or exceeding the species' physiological limits of sound production.

The acquisition of such baseline data (i.e. vocal repertoires and intraspecies vocal variation) would ideally consist of recordings of all behavioural contexts, life-history events, social interactions and developmental stages from several individuals of both sexes, possibly from several geographical

Table 1. List of commonly measured acoustic parameters, which are often reported very selectively in studies investigating VPL and/or characterizing species vocal repertoires. A comparative approach to the assessment of VPL would be greatly facilitated by the comprehensive reporting of as many of these parameters as possible throughout VPL studies. Exemplary references are given for each parameter. Note that not all parameters are equally well suited to characterize tonal and non-tonal vocalizations. We use the term ‘element’ here to indicate diverse kinds of vocalizations, such as calls, pulses, clicks and buzzes.

spectral parameters	amplitude characteristics
fundamental frequency (f_0) [3,10,14,24,26,29]	minimum, maximum level [26]
minimum, maximum f_0 [2,14,26,29]	envelope peaks [11]
start, end frequency of f_0 [24,26]	envelope skewness and kurtosis [24]
dominant harmonics [10]	envelope entropy [10,24]
peak frequency [3,14,26]	
spectral centroid [10,14,24,26]	temporal parameters
bandwidth (var. measures) [2,11,14,24,26,29]	element duration [10,11,14,24,26,29]
minimum, maximum frequency [26,29]	(inter-) element interval [11,26,29]
frequency modulation [10,24,26,29,30]	element rate (vocal activity) [11,26]
formant frequencies [2,3,10,24]	rhythm [31]
spectral envelope skewness and kurtosis [24]	time to maximum amplitude [29]
(spectral envelope) entropy/aperiodicity [10,14,24,26,30]	
	sequential characteristics
	order of elements [26,29,30]
	order of sequences [30]

locations. Acquiring such complete repertoires is extremely challenging, but they can be convincingly approximated. Long-term acoustic recordings of wild and captive animals often reveal the complexity of the vocal repertoire and allow educated guesses of the comprehensiveness of the recordings (e.g. for birds [23,24] and for mammals [25,26]). Such baseline knowledge of the vocal repertoire is crucial for the demonstration that a recorded call is indeed novel and did not pre-exist in the animal’s repertoire. The difficult evaluation of vocal novelty highlights not only the importance of clarifying which precise definition of VPL is applied but also the significance of a detailed reporting culture. The acquisition of reference data, such as call repertoires or baseline calls, should be a prerequisite for the assessment of the origin of newly arisen changes in vocal parameters.

(b) Proposed reporting of an acoustic parameter space

When suggesting a parameter space or a list of parameters to be reported, several things need to be said. Even though we want to stress the importance of comparability, there are always conditions under which the recording or reporting of parameters is not possible. For example, experimental conditions might prevent certain parameters from being recorded: artificial or natural background noise, constraints within the recording chain (limited sample rate, frequency range of the microphone/hydrophone, etc.), the acoustic character of the recording site (transmission loss, filtering characteristics, reverb), distance from the sound source, and the observability of the animal under investigation. Considering these disclaimers, acoustic parameters must be reported as comprehensively as possible. Furthermore, the more parameters are reported the better, as this enables cross-species comparisons, thereby highlighting their usefulness. Several papers and guides have been published in the past, focusing on bioacoustics recording and reporting standards [27,28].

This and comparable literature should be consulted before designing bioacoustics experiments in order to demonstrate VPL. Here, we list a number of acoustic parameters, which are often reported in studies investigating VPL and would facilitate the comparison between studies if reported comprehensively and throughout all VPL studies (table 1). Given the diversity of vocalizations and their modes of production, reaching from nearly pure tones over complex tonal or ‘noisy’ vocal structures and rhythms to clicks, it is important to have in mind that not all parameters are equally well suited to characterize every vocalization. However, the acoustic parameters listed in table 1 are well suited to give at least a simple description of most types of vocalizations.

The software and algorithms commonly used to assess differences in vocalization parameters are varied but should in the best case result in comparable outcomes. Acoustic analyses are regularly conducted with commercial or open-source software (e.g. PRAAT [32], Raven [33], Avisoft [34], Sound Analysis Pro [20], SIGNAL (Engineering Design, Berkeley, USA), Audacity® [35], Lucinia [36]) or with self-written code in Matlab, Python, C++ or R. Some programmes provide toolboxes or packages, which already include perceptual algorithms, such as a built-in mel scale (e.g. the Matlab speech processing toolbox [37]). The drawback of these programs is that they are usually not ideal for large-scale batch processing of sound recordings. This is where clustering algorithms show their true capability; however, they require initial human validation and should not be trusted blindly. One such algorithm is VoICE [38], which aims to increase comparability across labs and species by unifying the analytical approach. Simply put, this algorithm scores acoustic similarity of vocal output and categorizes it into a hierarchical cluster tree. A different approach to classifying vocal output was described by Valente and colleagues [39], where the individual output of a juvenile zebra finch was segmented and then analysed by calculating the Euclidean distance to segments of the

tutor song. A match occurred when a similarity threshold was crossed and the results from this automated process matched the results from human pattern recognition analysis. Such similarity recognition algorithms are often originally applied in human speech recognition (e.g. Gaussian Mixture Models [40]), but based on the spectral content of vocalizations represented by Mel-frequency cepstral coefficients. A similar approach has also been successfully used to quantify variation in the structure of acoustic signals in bats [41,42] and other mammals (e.g. deer [43], odontocetes [44] and elephants [45]).

All programs and analysis algorithms have their benefits and downsides. It is important to thoroughly research the best method to analyse the data. For this, we need to keep in mind that the judgement of the suitability of a method can always be influenced by the past experiences of the judge. Just as the VPL signal receiver might have learned preferences for a specific type of signal, researchers might have a learned preference for analysis software. Furthermore, some approaches may be better for assessing sequential correlations, whereas others are better at spectral comparisons [21]. Cross comparison not just between species, but across platforms can identify robust acoustic signals, as well as nuances that may be specific to the analysis method. A re-evaluation of the used analysis method should be conducted with the aim of evaluating whether the used software is up-to-date and the most useful tool for the present study design.

(c) Proposed extended reporting of training conditions and curves

Aside from the spectro-temporal parameters measured in different experimental conditions and the employed analysis software or strategy, several other factors require consideration and should be reported in order to facilitate VPL comparisons and the relative positioning of species on any proposed VPL continuum. For example, the time needed to achieve a certain degree of similarity should be considered. If limited training time for a few days results in the same degree of imitation of a specific template that is reached by a different species only through constant training for weeks, this should also be reported. The same is true for the conditions of the animal housing outside of the experiment. Isolated animals might have a higher motivation to participate in 'enriching' experiments, while animals that are kept with conspecifics might take a longer time to internalize a task or change. The comprehensive and comparative assessment of a species' VPL capacity expands even further and includes the number of trained individuals. Non-reported preliminary studies selecting for good learners blur the actual evaluation of the number of individuals willing and/or able to learn the task. Reporting the overall number of trained individuals does not indicate the species capacity for VPL, but would help to assess the species' overall willingness to learn the VPL task. This could help to select suitable model species and to make decisions about required sample sizes.

3. The behavioural decision-making space

(a) Behavioural evaluation of vocal production learning

When assessing VPL, bioacoustics measurements often give a result in the form of a test of statistical significance. Such a test might indicate that there is a significant difference between

initial and trained vocalizations, and yet the meaning or behavioural relevance of the vocalization might still be the same for the species. Conversely, a statistical test might indicate that a small difference in acoustic measurements is not significant, and yet this small difference has a marked effect on the communicative function of the vocalizations. The degree of change in vocalizations, which we observe with quantitative measurement methods, is thus not necessarily a good approximation of the biological relevance of a change in the signal. For example, Nowicki and colleagues showed that female song sparrows responded significantly more to songs that had been learned slightly better, demonstrating that variation in learning abilities plays a significant functional role in sexual selection [46].

In order to assess the biological importance of a learned vocal modification, external validation is helpful and often necessary. This external validation can be done using, for example, behavioural assays or neuronal representation of change and it presents a functional assessment of VPL. There are several approaches that can be taken if an external validation of VPL is desired. These different approaches depend entirely on the aim of the study. In the most common case, the quality of imitation can be demonstrated by reaching a level of difference too low to allow discrimination by the receiver. This means that vocal imitation can be demonstrated by a behavioural test for acoustic indistinguishability. Another way to demonstrate VPL would be to imitate a heterospecific vocalization well enough to convey a meaningful message. This is the case in the study mentioned above, in which humans discriminate words vocalized by an elephant [2]. The simplest idea would be to train animals to discriminate between novel and known vocalizations. This has been done among others in zebra finches [47], in starlings [48] and in swamp sparrows [49], which discriminated between different conspecific syllables or songs.

Comparable experiments with wild animals are also conceivable. For example, if an animal were trained to learn to imitate the dialect of a foreign population, the success of this VPL experiment could be quantified by, for example, a phonotaxis-based behavioural experiment. Similar approaches have been used in discrimination experiments in the past. Playback experiments with greater spear-nosed bats indicated that these bats could discriminate between individuals from different caves as well as between individuals from their own social group and foreign groups based solely on call structure [50]. Using a similar experimental design, Knörnschild and colleagues [51] were able to demonstrate that playbacks of local territorial song of male sac-winged bats attracted females more strongly than foreign territorial songs. Both examples show that the assessment of vocal divergence is possible and feasible in these species. Other acoustic discrimination experiments are also conceivable for the behavioural assessment of VPL: a study in nightingales showed that males increased their sound level significantly when presented with playbacks of conspecific rivals, however showed only little changes to their sound level output when the playback was of a heterospecific bird [52]. Another study showing that the meaning or information of a vocalization can be extracted, even if emitted by a heterospecific, was done in marmots and ground squirrels [53], thus demonstrating that the behavioural relevance of a vocalization can be maintained even if the acoustic parameters vary significantly between emitters. The opposite was shown for isolated songs of song

sparrows, which resemble natural conspecific song in several aspects, but do not generate the same response in conspecifics [54]. The aforementioned experiments relied on reactions of freely behaving wild animals, which is only feasible in studies where observation is possible, a luxury often not available when working with, for example, marine or solitary animals. Therefore, as mentioned above, direct human validation of the emitted vocalizations is often still required.

Another possibility to evaluate the degree of vocal imitation and to report it in a comparable way is to assess the observed vocal change in relation to the perception of the signal (e.g. auditory filters of the receiver of the newly learned signal). There is no biological need to imitate a signal perfectly if the sensory system of the receiver is not capable of discerning the minute differences between the target and imitated signal. For validation of the discriminability of newly learned vocalization and existing ones, models of the auditory periphery for humans and animals [55–57] should be used. These models create a spectro-temporal representation of sounds as a function of tonotopic frequency and time, the so-called auditory spectrogram [55]. This approach is typically used to describe physiological mechanisms underlying the perception of certain acoustic parameters and to explain human and animal performance in psychophysical discrimination tasks. Models of the auditory periphery typically employ implementations of the middle ear and cochlear functionalities (e.g. middle ear transfer characteristics, frequency-to-place conversion, nonlinear transformations of the organ of Corti, temporal integration) combined with an implemented decision device acting as an optimal detector based on the principles of signal detection theory [58,59]. For example, these models have been successfully used to describe behavioural discrimination performance in bats [57,60,61] and could well be used to evaluate VPL from a receivers' perspective. However, it should be kept in mind that such models are often based on approximations as experimentally derived information about model parameters is not available for many animal species. An open-source toolbox including numerous models for different stages of the (human) auditory system is available (<http://amtoolbox.sourceforge.net>), which might serve as a basis for applications in other species, too.

The judgement of imitation quality by a conspecific, heterospecific or perceptual model is an important criterion to assess the biological function of VPL and is critical for the investigation of the evolutionary origins of this trait. Therefore, a behavioural approach to quantify the extent and function of VPL in a species should always be considered as an important supporting experiment.

(b) Neuronal validation of vocal production learning via the receiver

The comparative approach will allow researchers to draw conclusions on the relative VPL capacity of species and thus help to uncover the biological basis of this trait [19]. A nuanced and comprehensive comparison of the VPL capacity of a preferably large number of species will ultimately allow insights into the neuronal basis of the human capacity for speech. VPL, just as speech, requires a high amount of auditory plasticity as it involves the use of auditory feedback to coordinate audio–vocal interactions while learning new vocalizations or while maintaining the stability of existing vocalizations [62]. Therefore, neural responses in the auditory

pathway should become selective to the new vocalizations to be learned, leading to an enhanced representation of new vocal elements [63]. Although we are aware that recording neural responses from auditory brain regions cannot be a standard procedure in every animal model used to study VPL, we want to briefly highlight the importance of the comparative study of the neuronal basis of VPL.

While the changes occurring in the neural network involved in vocal imitation and production have been studied in detail (for example, reviewed in [64,65]), the role of auditory forebrain areas in providing sensory feedback in VPL has not been studied in detail, but is receiving more and more attention [66]. In songbirds, neurons in the auditory forebrain were shown to encode information about the category of a vocalization but also about the identity of the emitter [67,68]. We here focus on the neural representation of newly learned vocalizations in forebrain areas involved in processing auditory input.

The first evidence for an emerging neuronal response selectivity for learned conspecific vocalizations in areas outside the song control system (a network of brain nuclei specialized for singing and song learning) came from the aforementioned study on adult starlings, which were trained to discriminate between conspecific songs [48]. Extracellular recordings (under anaesthesia) from neurons in the non-primary auditory forebrain region revealed a population of neurons showing a stronger response to familiar songs used in the training sessions when compared to novel songs. Thus, experience-driven plasticity seems to modify neural responses (and therefore the representation of conspecific vocalization) on the basis of the functional demands of song recognition. While these early results came from adult birds, a recent study showed specific changes in the neuronal representation of songs in juveniles being raised with heterospecific tutors [69]. They demonstrated that tuning for conspecific songs arises in the primary auditory cortical circuit of finches, as neurons showed stronger responses to conspecific songs than to songs of other species. Furthermore, this cortical representation could be shifted towards the songs of a tutor species in cross-fostering experiments. It was shown that the spectro-temporal tuning properties of neurons were altered to fit the spectro-temporal modulations of a learned song [69]. These findings support the indication that experience-dependent mechanisms might promote the alignment of auditory responses with the output of newly learned vocal motor-behaviour. Results from other studies on starlings hint in the same direction: male starlings raised without direct contact to adults not only failed to develop typical song classes but neurons in the caudomedial nidopallium (NCM, an analogue to mammalian secondary auditory cortex [70]) also failed to develop differential responses to different functional classes of song [71]. By contrast, differential NCM responses have been demonstrated in wild-caught starlings [72].

In mammals, evidence for changes in the sensory representation of species-specific communication due to VPL is still lacking. However, call-selective cortical neurons have been described in non-human primates [73,74]. Therefore, it can be assumed that similar changes as in songbirds can also be expected in auditory forebrain areas in mammalian species capable of VPL. However, it is important to note that changes in sensory representation in the auditory forebrain of birds and mammals can also occur independent of VPL, e.g. as a result of experience-dependent plasticity [75–78]. Where applicable, the neuronal validation of VPL through plastic changes in the

sensory representation of species-specific vocalization might, therefore, be an interesting additional tool to comprehensively investigate the capacity of VPL in a species. In addition to the investigation of neural activity and responses by the means of electrophysiological recordings, genetic methods can be used for the evaluation of acoustic signals. Specifically, immediate early gene expression has been used to identify active brain regions, e.g. during singing, song learning [79,80], and the perception of categorical different acoustic stimuli [81].

4. Conclusion and outlook

The aim of the guidelines provided here is to achieve a wide-reaching comparability in future reporting of findings concerning the VPL capacity of species. In order to achieve this, we highlight the importance of five factors influencing the unification of the VPL assessment: (i) comparison of vocal change to a well-established baseline, (ii) comprehensive reporting of acoustic parameters (not only significant ones), (iii) extended reporting of training conditions and durations, (iv) investigating VPL function via behavioural, perception-based experiments and (v) validation of findings on a neuronal level via the receiver. While the VPL capacity of a species can be successfully demonstrated without the inclusion of these factors, the comparison of cross-species VPL capacities is vitally dependent on our joint efforts to comprehensively study and report these factors.

A research culture in which a wide range of different acoustic parameters are routinely reported would allow us to draw conclusions about the VPL capacity of species independent of the current definition of vocal learning. Specifically, when comprehensive reporting of acoustic parameters is achieved, VPL capacity can easily be reassessed in cases of terminological or functional redefinition, e.g. due to the discovery of new mechanisms or forms of VPL. We want to

commend authors already following these reporting guidelines, and this paper should serve simply as a gentle reminder. However, the literature shows that this is not the case for the majority of reported studies, and we hope this paper can be used as a guideline for both study design and reporting of findings, therefore promoting future comparability between studies. In the future, approaches for the less human-centric evaluation of VPL will likely become more readily available. However, until we reach this golden age of easy, species-specific, perception-based evaluation algorithms, an important improvement of the current scientific practice would be attempts to evaluate VPL-related findings through the measurement of behavioural responses. In case such experiments are not feasible, the generation of well-designed potential follow-up experiments that would demonstrate the behavioural importance of findings would be beneficial to the field and increase future comparability between studies evaluating the VPL capacities of species.

Data accessibility. This article has no additional data.

Authors' contributions. All authors participated in the 'Unifying Vocal Learning' workshop in Leiden, 2019, laying the conceptual groundwork for the manuscript. E.Z.L. finalized the concept and wrote the outline of the manuscript. E.Z.L., U.F. and S.G.H. wrote large sections of the paper. All authors have worked on finalizing the paper and critically revised the manuscript.

Competing interests. We declare we have no competing interests.

Funding. We received no funding for this study.

Acknowledgements. The authors want to thank the participants of the Unifying Vocal Learning workshop in Leiden, 2019, for important, constructive group discussions and especially Katharina Riebel, Kazou Okanoya, Robert F. Lachlan and Pedro Tiago Martins for valuable input on the topic of this manuscript. Furthermore, we are grateful to Ofer Tchernikovski, Katharina Riebel and Vincent Janik for constructive feedback on earlier versions of this manuscript and Mark D. Scherz for critical proof-reading. Last but not least, we want to thank Frederic Theunissen and five anonymous reviewers for helpful comments and suggestions during the review process.

References

- 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)
- 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, Janik VM. 2019 Formant modification through vocal production learning in gray seals. *Curr. Biol.* **29**, 2244–2249. (doi:10.1016/j.cub.2019.05.071)
- Klatt DH, Stefanski RA. 1974 How does a mynah bird imitate human speech? *J. Acoust. Soc. Am.* **55**, 822–832. (doi:10.1121/1.1914607)
- Pepperberg IM. 2010 Vocal learning in grey parrots: a brief review of perception, production, and cross-species comparisons. *Brain Lang.* **11**, 81–91. (doi:10.1016/j.bandl.2009.11.002)
- Murayama T, Iijima S, Katsumata H, Kazutoshi A. 2014 Vocal imitation of human speech, synthetic sounds and beluga sounds, by a beluga (*Delphinapterus leucas*). *Int. J. Comp. Psychol.* **27**, 369–384. (doi:10.46867/ijcp.2014.27.03.10)
- Ridgway S, Carder D, Jeffries M, Todd M. 2012 Spontaneous human speech mimicry by a cetacean. *Curr. Biol.* **22**, 860–861. (doi:10.1016/j.cub.2012.08.044)
- Abramson JZ, Hernández-Lloreda MV, García L, Colmenares F, Aboitiz F, Call J. 2018 Imitation of novel conspecific and human speech sounds in the killer whale (*Orcinus orca*). *Proc. R. Soc. B* **285**, 20172171. (doi:10.1098/rspb.2017.2171)
- Vernes SC, Wilkinson GS. 2019 Behaviour, biology and evolution of vocal learning in bats. *Phil. Trans. R. Soc. B* **375**, 20190061. (doi:10.1098/rstb.2019.0061)
- Prat Y, Taub M, Yovel Y. 2015 Vocal learning in a social mammal: demonstrated by isolation and playback experiments in bats. *Sci. Adv.* **1**, e1500019. (doi:10.1126/sciadv.1500019)
- Boughman JW. 1998 Vocal learning by greater spear-nosed bats. *Proc. R. Soc. Lond. B* **265**, 227–233. (doi:10.1098/rspb.1998.0286)
- Esser K-H. 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)
- Prat Y, Azoulay L, Dor R, Yovel Y. 2017 Crowd vocal learning induces vocal dialects in bats: playback of conspecifics shapes fundamental frequency usage by pups. *PLoS Biol.* **15**, 1–14. (doi:10.1371/journal.pbio.2002556)
- Genzel D, Desai J, Paras YM. 2019 Long-term and persistent vocal plasticity in adult bats. *Nat. Commun.* **10**, 3372. (doi:10.1038/s41467-019-11350-2)
- Lattenkamp EZ, Vernes SC, Wiegrebte L. 2020 Vocal production learning in the pale spear-nosed bat, *Phyllostomus discolor*. *Biol. Lett.* **16**, 20190928. (doi:10.1098/rsbl.2019.0928)
- Jones G, Ransome RD. 1993 Echolocation calls of bats are influenced by maternal effects and change over a lifetime. *Proc. R. Soc. Lond. B* **252**, 125–128. (doi:10.1098/rspb.1993.0055)
- Knörnschild M, Behr O, von Helversen O. 2006 Babbling behavior in the sac-winged bat

- (*Scopteryx bilineata*). *Sci. Nat.* **93**, 451–454. (doi:10.1007/s00114-006-0127-9)
18. Knörnschild M, Nagy M, Metz M, Mayer F, von Helversen O. 2010 Complex vocal imitation during ontogeny in a bat. *Biol. Lett.* **6**, 156–159. (doi:10.1098/rsbl.2009.0685)
 19. 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)
 20. Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP. 2000 A procedure for an automated measurement of song similarity. *Anim. Behav.* **59**, 1167–1176. (doi:10.1006/anbe.1999.1416)
 21. Mandelblat-Cerf Y, Fee MS. 2014 An automated procedure for evaluating song imitation. *PLoS ONE* **9**, e96484. (doi:10.1371/journal.pone.0096484)
 22. Tyack PL. 2019 A taxonomy for vocal learning. *Phil. Trans. R. Soc. B* **375**, 20180406. (doi:10.1098/rstb.2018.0406)
 23. Adret-Hausberger M. 1989 The species-repertoire of whistled songs in the European starling: species-specific characteristics and variability. *Bioacoustics* **2**, 137–162. (doi:10.1080/09524622.1989.9753123)
 24. Elie JE, Theunissen FE. 2016 The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Anim. Cogn.* **19**, 285–315. (doi:10.1007/s10071-015-0933-6)
 25. Mumm CAS, Knörnschild M. 2014 The vocal repertoire of adult and neonate giant otters (*Pteronura brasiliensis*). *PLoS ONE* **9**, e112562. (doi:10.1371/journal.pone.0112562)
 26. Lattenkamp EZ, Shields SM, Schutte M, Richter J, Linnenschmidt M, Vernes SC, Wiegrebe L. 2019 The vocal repertoire of pale spear-nosed bats in a social roosting context. *Front. Ecol. Environ.* **7**, 16. (doi:10.3389/fevo.2019.00116)
 27. Beeman K. 1998 Digital signal analysis, editing, and synthesis. In *Animal acoustic communication* (eds SL Hopp, MJ Owren, CS Evans), pp. 59–103. Berlin, Germany: Springer.
 28. Köhler J *et al.* 2017 The use of bioacoustics in anuran taxonomy: theory, terminology, methods and recommendations for best practice. *Zootaxa* **4251**, 1–124. (doi:10.11646/zootaxa.4251.1.1)
 29. Knörnschild M, Nagy M, Metz M, Mayer F, von Helversen O. 2012 Learned vocal group signatures in the polygynous bat *Scopteryx bilineata*. *Anim. Behav.* **84**, 761–769. (doi:10.1016/j.anbehav.2012.06.029)
 30. Haesler S, Rochefort C, Georgi B, Licznernski P, Osten P, Scharff C. 2007 Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. *PLoS Biol.* **5**, 2885–2897. (doi:10.1371/journal.pbio.0050321)
 31. Roeske TC, Tchernichovski O, Poeppel D, Jacoby N. 2020 Categorical rhythms are shared between songbirds and humans. *Curr. Biol.* **30**, 1–12. (doi:10.1016/j.cub.2020.06.072)
 32. Boersma P, Weenink D. 2001 PRAAT, a system for doing phonetics by computer. *Glott Int.* **5**, 341–345.
 33. Center for Conservation Bioacoustics. 2019 *Raven Pro: interactive sound analysis software (version 1.6.1)* [Computer software]. Ithaca, NY: The Cornell Lab of Ornithology. See <http://ravensoundssoftware.com/>.
 34. Specht R. 2002 *Avisoft-saslab pro: sound analysis and synthesis laboratory* [Computer software]. Avisoft Bioacoustics. Berlin, Germany: avisoft.com.
 35. Audacity Team. 2020 *Audacity®: free audio editor and recorder*. See <https://www.audacityteam.org/>.
 36. [Computer software]. See <http://github.com/rflachlan/luscinia>.
 37. Ankitkumar C. 2020 *Speech processing toolbox*. See <https://www.mathworks.com/matlabcentral/fileexchange/39015-speech-processing-toolbox>.
 38. Burkett Z, Day N, Peñagarikano O, Geschwind DH, White SA. 2015 VoICE: a semi-automated pipeline for standardizing vocal analysis across models. *Sci. Rep.* **5**, 10237. (doi:10.1038/srep10237)
 39. Valente D, Wang H, Andrews P, Mitra PP, Saar S, Tchernichovski O, Golani I, Benjamini Y. 2007 Characterizing animal behavior through audio and video signal processing. *IEEE Multimedia* **14**, 32–41. (doi:10.1109/MMUL.2007.71)
 40. Reynolds DA, Quatieri TF, Dunn RB. 2000 Speaker verification using adapted Gaussian mixture models. *Digit. Signal Process.* **10**, 19–41. (doi:10.1006/dspr.1999.0361)
 41. Prat Y, Taub M, Yovel Y. 2016 Everyday bat vocalizations contain information about emitter, addressee, context, and behavior. *Sci. Rep.* **6**, 39419. (doi:10.1038/srep39419)
 42. Araya-Salas M, Hernández-Pinsón HA, Rojas N, Chaverri G. 2020 Ontogeny of an interactive call-and-response system in Spix's disc-winged bats. *Anim. Behav.* **166**, 233–245. (doi:10.1016/j.anbehav.2020.05.018)
 43. Reby D, André-Obrecht R, Galinier A, Farinas J, Cargnelutti B. 2006 Cepstral coefficients and hidden Markov models reveal idiosyncratic voice characteristics in red deer (*Cervus elaphus*) stags. *J. Acoust. Soc. Am.* **120**, 4080–4089. (doi:10.1121/1.2358006)
 44. Roch MA, Soldevilla MS, Burtenshaw JC, Henderson EE, Hildebrand JA. 2007 Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California. *J. Acoust. Soc. Am.* **121**, 1737–1748. (doi:10.1121/1.2400663)
 45. Clemins PJ, Johnson MT. 2013 Automatic type classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations. *J. Acoust. Soc. Am.* **113**, 2306. (doi:10.1121/1.4780702)
 46. Nowicki S, Searcy WA, Peters S. 2002 Quality of song learning affects female response to male bird song. *Proc. R. Soc. Lond. B* **269**, 1949–1954. (doi:10.1098/rspb.2002.2124)
 47. Chen J, van Rossum D, ten Cate C. 2015 Artificial grammar learning in zebra finches and human adults: XYX versus XXY. *Anim. Cogn.* **18**, 151–164. (doi:10.1007/s10071-014-0786-4)
 48. Gentner TQ, Margoliash D. 2003 Neuronal populations and single cells representing learned auditory objects. *Nature* **424**, 669–674. (doi:10.1038/nature01731)
 49. 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)
 50. Boughman JW, Wilkinson GS. 1998 Greater spear-nosed bats discriminate group mates by vocalizations. *Anim. Behav.* **55**, 1717–1732. (doi:10.1006/anbe.1997.0721)
 51. Knörnschild M, Blüml S, Steidl P, Eckenweber M, Nagy M. 2017 Bat songs as acoustic beacons—male territorial songs attract dispersing females. *Sci. Rep.* **7**, 13918. (doi:10.1038/s41598-017-14434-5)
 52. Brumm H, Todt D. 2004 Male–male vocal interactions and the adjustment of song amplitude in a territorial bird. *Anim. Behav.* **67**, 281–286. (doi:10.1016/j.anbehav.2003.06.006)
 53. Shriner WM. 1998 Yellow-bellied marmot and golden-mantled ground squirrel responses to heterospecific alarm calls. *Anim. Behav.* **55**, 529–536. (doi:10.1006/anbe.1997.0623)
 54. Searcy WA, Marler P, Peters S. 1985 Songs of isolation-reared sparrows function in communication, but are significantly less effective than learned songs. *Behav. Ecol. Sociobiol.* **17**, 223–229. (doi:10.1007/BF00300140)
 55. Patterson RD, Allerhand MH, Giguere C. 1995 Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform. *J. Acoust. Soc. Am.* **98**, 1890–1894. (doi:10.1121/1.414456)
 56. Dau T, Püschel D, Kohlrausch A. 1996 A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure. *J. Acoust. Soc. Am.* **99**, 3615–3622. (doi:10.1121/1.414959)
 57. Wiegrebe L. 2008 An autocorrelation model of bat sonar. *Biol. Cybern.* **98**, 587–595. (doi:10.1007/s00422-008-0216-2)
 58. Green DM, Swets JA. 1966 *Signal detection theory and psychophysics*. New York, NY: Wiley.
 59. Dau T, Kollmeier B, Kohlrausch A. 1997 Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *J. Acoust. Soc. Am.* **102**, 2892–2905. (doi:10.1121/1.420344)
 60. Grunwald JE, Schornich S, Wiegrebe L. 2004 Classification of natural textures in echolocation. *Proc. Natl Acad. Sci. USA* **101**, 5670–5674. (doi:10.1073/pnas.0308029101)
 61. Firzlaff U, Schuchmann M, Grunwald JE, Schuller G, Wiegrebe L. 2007 Object-oriented echo perception and cortical representation in echolocating bats. *PLoS Biol.* **5**, e100. (doi:10.1371/journal.pbio.0050100)
 62. Brainard MS, Doupe AJ. 2000 Auditory feedback in learning and maintenance of vocal behaviour. *Nat. Rev. Neurosci.* **1**, 31–40. (doi:10.1038/35036205)
 63. Prather JF, Mooney R. 2004 Neural correlates of learned song in the avian forebrain: simultaneous representation of self and others. *Curr. Opin.*

- Neurobiol.* **14**, 496–502. (doi:10.1016/j.conb.2004.06.004)
64. Mooney R. 2009 Neurobiology of song learning. *Curr. Opin. Neurobiol.* **19**, 654–660. (doi:10.1016/j.conb.2009.10.004)
65. Sakata JT, Yazaki-Sugiyama Y. 2020 Neural circuits underlying vocal learning in songbirds. In *The neuroethology of birdsong*, vol. 71 (eds J Sakata, S Woolley, R Fay, A Popper). Berlin, Germany: Springer.
66. Elie JE, Hoffmann S, Dunning JL, Coleman MJ, Fortune ES, Prather JF. 2019 From perception to action: the role of auditory input in shaping vocal communication and social behaviors in birds. *Brain Behav. Evol.* **94**, 51–60. (doi:10.1159/000504380)
67. Elie JE, Theunissen FE. 2015 Meaning in the avian auditory cortex: neural representation of communication calls. *Eur. J. Neurosci.* **41**, 546–567. (doi: 10.1111/ejn.12812)
68. Elie JE, Theunissen FE. 2019 Invariant neural responses for sensory categories revealed by the time-varying information for communication calls. *PLoS Comp. Biol.* **15**, e1006698. (doi:10.1371/journal.pcbi.1006698)
69. Moore JM, Woolley SMN. 2019 Emergent tuning for learned vocalizations in auditory cortex. *Nat. Neurosci.* **22**, 1469–1476. (doi:10.1038/s41593-019-0458-4)
70. Terleph TA, Mello CV, Vicario DS. 2006 Auditory topography and temporal response dynamics of canary caudal telencephalon. *J. Neurophysiol.* **66**, 281–292. (doi:10.1002/neu.20219)
71. George I, Cousillas H, Richard JP, Hausberger M. 2008 A potential neural substrate for processing functional classes of complex acoustic signals. *PLoS ONE* **3**, e2203. (doi:10.1371/journal.pone.0002203)
72. George I, Alcaix S, Henry L, Richard JP, Cousillas H, Hausberger M. 2010 Neural correlates of experience-induced deficits in learned vocal communication. *PLoS ONE* **5**, e14347. (doi:10.1371/journal.pone.0014347)
73. Rauschecker JP, Tian B, Hauser M. 1995 Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* **268**, 111–114. (doi:10.1126/science.7701330)
74. Wang X, Kadia SC. 2001 Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J. Neurophysiol.* **86**, 2616–2620. (doi:10.1152/jn.2001.86.5.2616)
75. Thompson JV, Gentner TQ. 2010 Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. *J. Neurophysiol.* **103**, 1785–1797. (doi:10.1152/jn.00885.2009)
76. Thompson JV, Jeanne JM, Gentner TQ. 2013 Local inhibition modulates learning-dependent song encoding in the songbird auditory cortex. *J. Neurophysiol.* **109**, 721–733. (doi:10.1152/jn.00262.2012)
77. Fritz JB, Elhilali M, Shamma SA. 2007 Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J. Neurophysiol.* **98**, 2337–2346. (doi:10.1152/jn.00552.2007)
78. Gao E, Suga N. 2000 Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: role of the corticofugal system. *Proc. Natl Acad. Sci. USA* **97**, 8081–8086. (doi:10.1073/pnas.97.14.8081)
79. Feenders G, Liedvogel M, Rivas M, Zapka M, Horita H, Hara E, Wada K, Mouritsen H, Jarvis ED. 2008 Molecular mapping of movement-associated areas in the avian brain: a motor theory for vocal learning origin. *PLoS ONE* **3**, e1768. (10.1371/journal.pone.0001768)
80. Gobes SMH, 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)
81. Van Ruijsevelt L, Chen Y, von Eugen K, Hamaide J, De Groof D, Verhoye M, Güntürkün O, Woolley SC, Van der Linden A. 2018 fMRI reveals a novel region for evaluating acoustic information for mate choice in a female songbird. *Curr. Biol.* **28**, 711–721. (doi:10.1016/j.cub.2018.01.048)