Convergent Behavioral Strategies and Neural Computations during Vocal Turn-taking across Diverse Species

Arkarup Banerjee[1,2] & Daniela Vallentin[3,4]

[1] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
[2] Cold Spring Harbor Laboratory School for Biological Sciences, Cold Spring Harbor, NY
[3] Max Planck Institute for Ornithology, 82319 Seewiesen, Germany
[4] Max Planck Institute for Biological Intelligence (in foundation), 82319 Seewiesen, Germany

Correspondence: abanerjee@cshl.edu; daniela.vallentin@orn.mpg.de

Abstract

Vocal exchanges between individuals are often coordinated in a temporally precise manner: one party is vocalizing while the other one is listening until the performance roles are switched. This vocal turn-taking behavior is widespread across the animal kingdom and thus provides an opportunity to study the neural circuit mechanisms from a comparative perspective. Although the physical prerequisites of the vocal tracts across animals can be different the behavioral outcome of turn taking is often similar with respect to vocal response timing and context dependent adaptation. Here we review behavioral strategies of vocal turn-taking in diverse animals. Further, we highlight recent advances in studying the neural circuit mechanisms underlying vocal production and perception.

Introduction

Adaptive behavior requires real-time adjustments in response to a rapidly changing environment. How the brain accomplishes this sensorimotor feat is one of the most fascinating and puzzling questions in neuroscience. Although examples of animals engaged in adaptive behavior abound in the natural world, understanding the neural circuit mechanisms require experimentally tractable model systems and behavioral paradigms. For many species, including humans, one central adaptive behavior for social interactions is vocal communication. Such communication can arguably be unidirectional, but here we restrict ourselves to phenomena requiring bi-directional information flow between the sender and the receiver. Such vocal turn-taking requires the perception of auditory signals from the sender to generate appropriate vocal motor outputs making it ideally suited to study dynamic sensorimotor integration.

Using sounds to communicate is widespread in nature; from chirping birds [1] to duetting lemurs [2], a large number of species use vocalizations to cooperate as well as to compete. These vocal

39 interactions often follow specific temporal patterns whereby two individuals exchange
40 vocalizations in a coordinated manner avoiding overlaps and taking turns. Humans engaged in
41 conversation, for example, take rapid turns to go back and forth (turn-taking) between listening
42 and responding – a feat most of us tend to perform effortlessly, but which breaks down during
43 communicative disorders [3].
44
45 In this review, we will first highlight behavioral evidence for vocal turn-taking across the animal
46 kingdom, discussing its ubiquity, its ecological function and how it may represent a common
47 framework to study hitherto disparate phenomenon. Next, we will discuss recent developments in
48 our understanding of neural circuit mechanisms for vocal communication taking examples from a
49 few different species. Ultimately, we hope to convey that studying vocal turn-taking, at the
50 intersection of ethology and systems neuroscience, represents a fruitful path forward in our quest
51 to better understand the neurobiology of sensorimotor integration underlying adaptive behaviors.
52
53 **Vocal turn-taking: a unifying framework to study diverse phenomenon**
54
55 **a) Vocal turn- taking is widespread across the animal kingdom**
56 Spoken conversations between humans involve alternating exchange of vocalizations that tend to
57 overlap minimally [4]. This vocal turn-taking behavior not only requires fast sensory perception
58 of the sender's vocal output but also the precise control of one's own vocal onset. During these
59 interactions, participants simultaneously plan upcoming vocalizations while listening in order to
60 respond as early as possible [5]. In humans, response times are typically about 250 ms (Fig. 1 A),
61 although average latency can vary across linguistic cultures [6]. Recent work has shown that
62 speech planning activity can be localized to a frontotemporal brain network distinct from those
63 involved in speech perception and production *per se* [7]. Due to the necessity of coordination in
64 turn-based speech, and its early onset in ontogeny, vocal interaction has been proposed as a core
65 component for language acquisition [8].
66
67 Vocal turn-taking in mammals is not restricted to humans; non-human primates as well as some
68 rodent species have been shown to interact with each other by coordinating their vocalizations.
69 Common marmosets (*Callithrix jacchus)* live in large groups of 3-15 individuals and communicate
70 with each other by performing 13 different calls which are used in specific contexts to indicate
71 food, distance or excitement [9]. Specifically, phee calls (example of a distance call) are
72 preferentially used during vocal turn taking [10, 11]. In the family of mongooses, meerkats
73 (*Suricata suricatta*), have been observed to communicate with each other using coordinated vocal
74 interactions on sunny mornings during the cold season [12]. Timing of vocalizations could signal
75 individual identity, internal state, or estimated physical distance (especially during phee calls in
76 marmosets when individuals are out of site). Conveniently, marmosets engage in vocal interactions
77 with playbacks in a laboratory setting [11], which is ideal to systematically explore how context
78 affects vocal turn-taking. Lab mice or rats, however, have not been shown to vocally interact with

79 conspecifics by taking turns. Notably, the Alston's singing mouse has recently been found to
80 exhibit vocal turn taking behavior [13]**(for details see Section below).
81
82 A spectacular elaboration of turn-taking behavior is observed in avian species. Songbirds have
83 been extensively studied for their exquisite vocal turn-taking performances. Their behavior ranges
84 from the exchange of simple innate calls in zebra finches (*Taeniopygia guttata*) [14] to complex
85 interactions with hundreds of learned songs in nightingales (*Luscinia megarynchos*) [15]. Many
86 tropical bird species perform duets in opposite-sex pairs during which they sing cooperatively to
87 defend a territory or reinforce the pair bond [16]. Some duetting styles (i.e., antiphonal duets) are
88 characterized by synchronized time- and pattern- specific responses between partners, in which
89 each bird rapidly adjusts its vocalizations over the course of an interaction depending on the
90 partner. During duets, plain tailed wrens, for instance, time their song more accurately compared
91 to solo singing. Moreover, males sing louder with a female suggesting context-dependent song
92 adjustments [17, 18]. Recent studies on white-browed sparrow weavers further indicate that
93 duetting birds accurately synchronize their vocal output to avoid overlaps and optimize their
94 alternating turn-taking behavior, which sometimes even follows pair specific rules [19-21]**.
95
96 Notably, songbirds with a very limited vocal repertoire also exhibit precise vocal turn-taking
97 behavior. Throughout the day, a male zebra finch switches between singing its stereotyped
98 courtship song and much shorter contact calls [14]. While contact calls are used to counter-sing
99 with vocal partners at timescales comparable to humans, songs are not. [22]**. In addition, zebra
100 finches are capable of adapting their response strategy depending on the social context by flexibly
101 adjusting the timing of their calls in relation to a vocal partner. This ability might be relevant for
102 other contexts that go beyond mate attraction or territorial defense; during group flights, for
103 example, individual birds may potentially coordinate their calls indicating their position, velocity
104 or directional changes.
105
106 The phenomenon of turn-taking is truly widespread in the animal kingdom. Mammalian lineages,
107 such as bats and dolphins, which inhabit terrestrial and aquatic niches respectively, participate in
108 vocal turn-taking as well. For example, adult white-winged Vampire bats [23] use antiphonal
109 calling to communicate outside of their roost. Accumulating evidence suggests that multiple bat
110 species on the one hand learn their vocal repertoire [24, 25] and on the other hand use it in a more
111 complex manner than previously thought [26]. Dolphins produce characteristic whistles during
112 coordinated vocal interactions, which seem to facilitate individual recognition and maintenance of
113 group cohesion [27]. Going even beyond vertebrates, turn-taking behavior is also prevalent in
114 insects [28]. One prominent example is *Drosophila virilis*, a species of fruit fly that engages in
115 acoustic duetting [29]. Further exploration of vocal-turn taking behavior in other species (e.g. the
116 fish *Danionella cerebrum* which can vocalize [30]) will reveal whether this phenomenon is even
117 more ubiquitous than previously appreciated.
118

119

b) The potential function of vocal turn-taking

120 The ecological significance of vocal turn-taking behavior ranges from reproductive competition
121 and conflict to sexual selection strategies. To avoid overlap with a communication partner implies
122 that vocal signal should not be masked but clearly conveyed. This is useful in the framework of
123 territorial defense when the aim is to defeat a vocal sparring partner or, alternatively during mating
124 when females are supposed to be attracted by two individuals. Targeted overlap of a
125 communication signal can also be interpreted as a sign of aggression to perturb the performance
126 of the vocal partner. However, whether vocal overlap is a sign of aggression is debated [31, 32]
127 and further studies are needed to clarify this issue.

129

130 The ability to coordinate vocalizations in an interspersed manner precedes spoken language
131 developmentally and evolutionarily, extending to other species ranging from non-human primates
132 to birds and anurans [33]. Antiphonal advertisement calls of frog have been well studied and we
133 would like to refer the reader to Wells 1977 [34] for an extensive review. In all cases, vocal
134 interactions generally involve perceiving relevant acoustic signals and initiating exact motor
135 commands to generate an appropriate vocal reply. This social form of sensorimotor coordination
136 reduces acoustic overlap, thereby ensuring that signals are accurately transmitted and can be
137 correctly detected. Signaler and listener can respond to each other's vocalizations with variable
138 delay, together with other adjusted structural vocal parameters for communicating their current
139 state [35]. The fundamental principles of vocal turn-taking can only be fully understood when both
140 participants and their associated internal states are taken into account.

141

142 All vocal turn-taking species partition acoustic space to maximize information transfer. For
143 example, many anurans are highly vocal and communal social interactions are characterized by
144 non-overlapping vocalizations of multiple individuals. However, this strategy can also be changed
145 to achieve maximal overlap which results in synchronous vocal outputs. It is hypothesized that the
146 switch between these strategies might be an epiphenomenon to reset individual's signal output
147 upon hearing a neighbor's signal [36]. Interestingly, when multiple species habitat the same niche,
148 they vocalize at different times of the day to minimize acoustic overlap. For instance, frog species
149 within a community take turns with other species by producing their advertisement calls only at
150 specific times during the day while the other species remains quiet [37]. Although not strictly vocal
151 turn-taking between two conspecifics as defined above, here multiple species are taking vocal turns
152 sequentially. This provides further evidence that communication evolves to minimize acoustic
153 overlap.

154

155 Interestingly, in most animals only certain vocal types are reserved for turn-taking, whereas the
156 rest are used in the solo context. For example, in zebra finches, for vocal usage depends upon
157 specific contexts. During social bonding, they use short calls but switch to producing stereotyped
158 songs (in a unidirectional manner) while attracting females. Similarly, marmosets specifically use

159 phee calls for antiphonal interactions but have a larger repertoire of vocalizations for signaling
160 food or emotional state. One potential benefit of having specialized call types for vocal turn-taking
161 might be to unambiguously signal a conspecific with an explicit expectation of a response in return.
162
163 Vertebrates have developed numerous ways of producing sounds for communication. For example,
164 some fish drum on their swim bladder [38] or stridulate their pectoral fins [39]; amphibians,
165 reptiles, and mammals have a larynx that functions as a voice box during vocalizations [40] ; and
166 birds have a syrinx with membranes that vibrate during singing [41]. The diversity of vocal organs
167 in vertebrates is partly mirrored by different neural structures controlling the vocal organ, but there
168 are also similarities among species. Recent findings suggest that physiological and functional
169 properties of neuronal circuits in birds and mammals are comparable, and analogous circuits can
170 be identified in endbrain structures [42-46] as well as at the brainstem level in all vertebrates.
171 Taken together, all vocal turn-taking species achieve the same behavior with a slightly different
172 toolkit. In addition, since different species can either share communication strategies for vocal
173 interactions or use distinct strategies, a comparative approach to study this behavior will provide
174 insights into the canonical circuit functions and identify specialized ones underlying vocal control
175 and production, as well as audio-vocal integration processes.
176
177
178 **What are the neural principles underlying vocal turn-taking in mammals?**
179
180 The neural processes underlying vocal communication behaviors are quite complex. It requires the
181 integration of auditory inputs, planning, and generation of appropriate motor commands to move
182 the vocal muscles, all within a fraction of a second. Since, lesions in different brain regions can
183 cause specific deficits (e.g., sensory or motor aphasias), we have a parts-list of different brain
184 regions mediating vocal communication. However, we do not understand the neural circuit-level
185 mechanisms of vocal communication, despite its relevance to many communication disorders.
186
187 In mammals, species-typical sounds are generated by the primary "vocal motor network",
188 consisting of evolutionarily conserved brain areas in the midbrain and brainstem [45]. Based on
189 human brain lesions, gene expression profiles and neurophysiology evidence in primates, a second
190 frontal/motor "volitional control system" has been proposed [47]. Such cortical control over vocal
191 production circuits is thought to render flexibility to vocal behaviors, such as during conversations
192 or for modifying vocal usage based on context [48]. A key player of the primary "vocal motor
193 network" is the periaqueductal grey (PAG). PAG lesions result in mutism and stimulation, either
194 electrical or chemical, generates species-typical vocalizations in monkeys [49]. A recent study in
195 laboratory mice has revealed the central importance of PAG in controlling ultrasonic vocalizations
196 (USV). The first step was to use a clever genetic strategy to selectively label PAG neurons that are
197 transiently activated in relation to USVs, referred to as the PAG-USV neurons [50]**. Blocking
198 neurotransmission of the PAG-USV neurons using tetanus toxin light-chain reduced USV
199 production and conversely, chemical or optogenetic activation led to more USVs in the absence of

social cues. Importantly, optogenetic stimulation evoked USVs were produced during the exhalation phase of the respiratory cycle and did not differ from natural USVs in acoustic features. Finally, the authors showed that selective activation of PAG-USV neurons that project to the nucleus retroambiguus in the brain-stem is sufficient to elicit USVs [50]**.

But what triggers or suppresses such vocal production circuits especially in a context dependent manner? In monkeys, stimulation of the anterior cingulate cortex (ACC) that projects directly to PAG, elicits species-specific vocalizations [51]. Other brain areas such as the hypothalamus and amygdala, when electrically stimulated also result in the production of vocalizations, suggesting that many regions upstream of PAG could influence context-specific vocalizations. Recent studies in laboratory mice have revealed the detailed circuit mechanism by which hypothalamus and amygdala control vocal production via the PAG [52, 53]. Optogenetic activation of PAG projecting amygdala neurons can transiently suppress ongoing USVs. In contrast, activation of GABAergic hypothalamic neurons (medial preoptic area) that project to PAG elicits USVs in the absence of social cues [52]. It has been shown that increased activity of the hypothalamic input to PAG (e.g., ESR1+ve neurons in the lateral preoptic area) can flexibly scale the duration and amplitude of the USVs [53]**. Together, these studies begin to define the core circuit elements of the "primary vocal motor network" wherein PAG-USV neurons control USV production via pattern generators in the brainstem and in turn, hypothalamic or amygdala inputs to PAG provide context-specific behavioral gating (Fig 1 D).

The ability to exert voluntary control over phonation is a crucial element of human speech. Does this have a phylogenetic precursor in other species? Behavioral evidence in favor of this idea comes from experiments demonstrating that macaque monkeys (*Macaca mulatta*) can be successfully trained to emit vocalizations (coo calls) in response to food [54] or to arbitrary visual stimuli [55]. Insights into neural circuit mechanisms for such voluntary control comes from neural (extracellular) recordings during volitional control of vocalizations. For example, macaque premotor cortex (PMv) contains vocalization-specific population of neurons that discharge either immediately before or during vocal onset, showing a preference for conditioned (voluntary) vocalizations comparted to spontaneous ones [55]. Similar results have been observed in the ventrolateral prefrontal cortex (vlPFC), the putative Broca's area analog in monkeys. Macaque monkeys were successfully trained on a Go No-Go task to vocalize in response to a Go visual stimulus or withhold their vocalizations in response to the No-Go visual stimulus [55]. Single-unit electrophysiology recordings during this paradigm revealed that vlPFC neurons discharged preferentially during cued vocalizations and not during arousal outside of training. Moreover, neural activity of specific neurons immediately before vocal onset was correlated with acoustic features such as the call duration. Taken together, these findings provide evidence for a dual network model of vocalizations, wherein a frontal/motor "volitional control system" involving multiple cortical areas (e.g., PFC and PMv) is capable of exerting voluntary control over the phylogenetically older "primary vocal motor network" described above (Fig. 1 D).

240

For turn-taking such as during a conversation, vocal production needs to be flexibly controlled by the auditory inputs from the vocal partner. What are the neural circuit mechanisms underlying such flexible auditory-motor coupling in the brain? For example, the New World marmoset monkeys participate in vocal exchanges. Consequently, marmosets have emerged as a prominent primate model to study vocal communication [10, 11, 56]. Marmosets have a large repertoire of vocalizations, including trills, twitters and phee calls that are used in distinct social settings [57]. Among these, the most extensively studied is the phee call – a long-range contact call, which the common marmoset (*Callithrix jacchus*) uses to participate in vocal turn-taking during cooperative vocal communication [10, 58] (Fig 1 B). This turn taking behavior is on the order of seconds (Fig 1 C) and thus, follows a slower time scale compared to human conversation. Phee calls are thought to encode a variety of social information such as caller sex and individual identity [59-61]. Recent studies have shown that marmosets are capable of interrupting and modulating ongoing vocalizations based on auditory inputs [44]. For example, auditory perturbation during on-going phee calls showed that animals interrupt vocalizations only at discreet time points, suggesting that each phee call is in fact composed of multiple vocal units [62]**. Further, it has been observed that marmosets rarely initiate vocalizations during the presentation of calls and therefore seem to inhibit calling to avoid interference [10, 11]

A few studies have measured neural activity in different cortical regions during perception and production of phee calls. Neurons in the marmoset primary auditory cortex respond to spectrotemporal acoustic pattern of the species-typical vocalizations [63]. In addition to these purely sensory responses, auditory cortex neurons show significant vocalization-induced modulation of firing rates during self-generated phee calls [64]. Most of these auditory cortical neurons show a suppression of neural activity that starts prior to the onset of vocalizations, consistent with an efference copy mechanism thought to convey the predicted sensory consequences of self-generated movements [65]. In line with this model, these vocalization-suppressed neurons in auditory cortex were subsequently found to be sensitive to altered self-generated vocal feedback [66]. Moreover, neurons in auditory cortex are sensitive to compensatory vocal control in response to auditory feedback. Electrical stimulation of the same sites evokes rapid changes in vocal production [67]. Therefore, behavioral and physiological evidence suggests that auditory cortical activity plays a causal role in feedback-dependent vocal control [67].

What are the brain areas that control audio-vocal interactions in Marmosets? Ventral premotor cortex (vPMC) and prefrontal cortex (PFC) showed heightened immediate early gene expression during antiphonal conversations (Fig 1 D) [60, 68]. Subsequent electrophysiological recordings identified neurons in PFC and vPMC that show increased neural activity during vocal production [69]. Interestingly, these neurons are also active during spontaneously generated phee-calls in contrast to the vlPFC neurons reported in the macaques that are modulated only by cued-vocalizations [54, 55]. Further studies are needed to resolve whether this discrepancy arises due to

280 operant-conditioning in the case of the macaque experiments or it reflects species-specific
281 differences.
282
283 While these pioneering studies have identified the brain regions involved in primate turn-taking
284 behaviors, our understanding of neural circuit mechanisms underlying vocal interactions,
285 especially in mammals, remains quite rudimentary. While lab mice (*Mus musculus*) or rats (*Rattus*
286 *rattus*) produce ample ultrasonic vocalizations, they do not seem to take turns. Recently, another
287 mouse species, the Alston's singing mouse (*Scotinomys teguina*) has been found to not only
288 vocalize in the audible range for humans but also to exhibit vocal turn-taking with conspecifics
289 [70]. These rodents sing both spontaneous "solo" songs, as well as "duets". This counter-singing
290 behavior requires sub-second modification of motor outputs in response to auditory cues with a
291 temporal precision that resembles human conversation [13]. Importantly, vocalizations that are
292 relatively stable during solo singing become highly variable when the individual participates in
293 vocal turn-taking, demonstrating strong context dependence – a hallmark of social behaviors.
294 Furthermore, as a small rodent, it is readily amenable to many behavioral, electrophysiological,
295 and viral tools developed in recent decades. Therefore, a more detailed analysis of the underlying
296 turn-taking circuitry in mammals was conducted with the Alston's singing mice.
297
298 Using four complementary lines of evidence, Okobi, Banerjee et. al. defined a region of orofacial
299 motor cortex (OMC) that mediates flexible vocal interactions in the singing mouse. Intracortical
300 microstimulation (ICMS) was first used to map a functional hotspot located on the anterolateral
301 aspect of motor cortex, which was referred to as the orofacial motor cortex (OMC). Electrical
302 stimulation of OMC disrupted ongoing singing behavior. Additionally, mild focal cooling of OMC
303 slowed down the progression of the motor sequence by incorporating additional notes, resulting in
304 considerably longer songs. Furthermore, silencing OMC activity (via GABA-A agonist
305 muscimol), the authors showed that the animals could still sing their songs but could no longer
306 participate in vocal interactions. While previous studies have used immediate early gene
307 expression or electrophysiology to suggest cortical involvement in non-mammalian
308 communication, these results represent the first direct demonstration of motor cortical dependence
309 of vocal interactions in a non-primate mammalian species. These experiments demonstrated
310 hierarchical vocal motor control in a rodent [13], and extends the scope of the dual network model
311 of vocal production [47] that was previously thought to be evolutionarily restricted to the primate
312 lineage.
313
314 In summary, evidence from primate and rodents suggest that in all mammals, species-typical
315 sounds, like laughing or crying, which tend to be stereotyped and relatively inflexible, are
316 generated by the primary "vocal motor network". This network consists of vocal pattern generating
317 regions in the brainstem and operates under limbic control. There exists a second frontal/motor
318 "volitional control system", which exerts *cortical control* over this primary vocal motor network

[47]. The volitional control system effectively renders certain types of vocalizations, i.e., speech, more flexible.

Neural control of vocal turn-taking in birds

How is this 'vocal motor network' and the 'volitional control system' implemented in other animal species with a different brain architecture and network connectivity? Songbirds are among the most vocal animals and their brain is comprised of dedicated and well-defined pathways for auditory perception and vocal production [71] for which molecular and functional homologies to the mammalian cortex have been described [72]. Due to its anatomical location as an interface between the auditory pathway and the vocal motor pathway, the premotor nucleus HVC (proper name) is well positioned to play an important role during sensorimotor integration. Functionally, it has been shown that HVC shows stereotyped activity while a bird is singing [73] thus suggesting a role in patterned vocal signals, and likely in vocal turn-taking. Recent research by Benichov & Vallentin (2020) [22]** found that zebra finches tend to respond with preferred latencies to specific social partners. In the experiment, isochronous stack call playbacks were presented to an isolated bird. Different individuals responded with a specific stereotyped latency (198-322 ms). Besides, adding another bird to the set up resulted in one of the individuals responding later in time, showing that zebra finches might anticipate the calls of a vocal conspecific in order to adjust their own call timing, and ultimately avoid overlapping. The initial findings highlight the flexibility of the vocal system and its dependence on the social context. Following the behavioral characterization HVC was pharmacologically inactivated bilaterally. The response timing got reversibly impaired, the precision was lost, with birds no longer showing a preferred response time. Benichov et al. (2016) [74] observed similar results after RA lesion, as expected from being immediately downstream of HVC in the descending motor pathway. Finally, to identify the neural circuit mechanism within HVC related to call timing, Benichov & Vallentin (2020) performed intracellular recordings from HVC neurons while the birds were calling. Results showed cells with call-related premotor activity, call-related inhibition that preceded the premotor activity, and inhibitory interneurons transiently increasing their firing rate in relation to calling and then reducing it. Interestingly, the rise in interneuron activity preceded the premotor signal, suggesting a role for this type of neurons in regulating premotor cells in HVC, and therefore call timing. To test this hypothesis, inhibition within HVC was temporarily lifted by means of Gabazine application. This pharmacological manipulation resulted in a change in call timing i.e., birds were calling faster in response to the call playback by up to 100 ms. Thus, premotor neurons provide accuracy to vocal onsets whereas interneurons ensure that vocalizations are triggered at appropriate times. All in all, the findings underpin that HVC is necessary for call timing precision, thus clarifying the forebrain role in vocal turn-taking.

A similar role for inhibition during vocal turn-taking has been described in HVC of duetting wrens [75]. These animals counter sing in a rapid manner with each other. Just like in other songbirds, HVC neurons are active during song production but not during auditory input when these birds are

engaged in song production. To test the hypothesis whether inhibition within HVC suppresses auditory information during vocal interactions, the authors lifted inhibition by means of urethane anesthesia and showed that auditory-related activity reemerged. Taken together, inhibition might play a crucial role in withholding vocal production while listening to a communication partner.

Remarkably, in spite of millions of years of divergence between birds and the singing mice, similar neural control circuitry underlies vocal turn-taking behavior. This underscores the importance of the comparative neuroethology approach to discover the canonical neural principles of vocal turn-taking behavior in diverse species.

Conclusions and Outlook

In systems neuroscience, typical paradigms to study sensorimotor transformations involve training animals to associate arbitrary stimuli to specific actions. For example, in a Go No-Go task, a rodent may be trained to associate a sensory stimulus with a specific movement and also to withhold the movement in response to a different sensory stimulus (Fig 2 A). Such paradigms allow experimenters to precisely control sensory stimuli as well as measure behavioral outputs with great precision and have taught us a lot about neural computations and circuit-mechanisms underlying such sensorimotor behaviors.

There is growing emphasis on using naturalistic behaviors in neuroscience [76-78]. Vocal communication, especially turn-taking behaviors described above goes beyond 'naturalistic' – they represent tractable natural and self-guided behaviors ideally suited to study neural computations underlying sensorimotor integration. Since vocal partners take turns and avoid jamming each other as described above, this behavior can be conveniently delineated into sensory, delay and motor epochs, a feature typically engineered in many operant conditioning tasks (Fig 2). Therefore, understanding the neural basis of vocal turn-taking behaviors in diverse species offers a potentially paradigm-changing approach for investigating the substrates of ethologically relevant perception, cognition and action in the laboratory (Fig 2 B).

Finally, studying natural behaviors such as vocal turn-taking enables one to go beyond the 'proximate' question of how neural circuits function towards the 'ultimate' question of how neural circuits evolve (specified by genes and modified by learning). Comparing and contrasting neural circuit mechanisms across diverse species can help us to discover canonical algorithms underlying vocal turn-taking. Going forward, out of the plethora of questions that can be addressed while studying vocal turn-taking behavior, we would like to highlight the following three:

- Can a unifying signature of vocal turn-taking behavior be identified that holds true across animals? Or did different strategies evolve due to ecological needs?
- Are there common neural circuit motifs across animal species that govern the temporal basis of vocal turn-taking?

400   -   What are the relative contributions of cortical and subcortical areas to the execution of
401   vocal turn-taking during different contexts?

402   The spectacular progress in neural measurement and manipulation technologies in the last decade,
403   coupled with gene editing platforms such as CRISPR, can hopefully allow us to leverage the
404   diversity of natural behaviors in pursuit of generalizable principles in neuroscience.

405
406   **Highlights:**

407
408   **Benichov and Vallentin, 2020:** This study investigates the neural circuit mechanisms
409   underlying call interaction between zebra finches. Combining electrophysiological recordings of
410   identified cell types in the premotor nucleus HVC and subsequent targeted pharmacological
411   manipulations of these neurons the authors show that inhibition plays a key role in withholding
412   and executing vocal production at appropriate times.

413
414   **Okobi, Banerjee et al, 2019:** This study was the first to describe fast and flexible vocal
415   interactions in a neotropical species of rodent, the Alston's singing mice. Using a variety of neural
416   circuit perturbation experiments, the authors demonstrated robust and hierarchical motor cortical
417   control over vocal behaviors in a rodent species, previously thought to be restricted only to the
418   primate lineage.

419
420   **Tschida et al, 2019:** Using a clever genetic strategy, this study managed to specifically tag
421   ultrasonic vocalization sensitive neurons in the periaqueductal grey (PAG). Using bi-directional
422   causal manipulations, the authors demonstrated the central importance of PAG in gating USV
423   production in lab mice.

424
425   **Hoffmann et al, 2019**: The authors investigated the neural basis of vocal duetting behavior in
426   songbird pairs ranging freely in their natural habitats. Wireless technology enabled simultaneous
427   recordings of individual vocalizations and multiunit vocal premotor activity. They found evidence
428   for inter-individual synchrony in the premotor activity in the nucleus HVC between duetting song
429   birds.

430
431   **Pomberger et al, 2018:** The authors investigate whether marmoset monkeys are able to produce
432   vocalizations with variable durations dependent on external sensory perturbations. They used a
433   noise paradigm to perturb the marmosets during phee call production and find that these calls can
434   be segmented in defined components indicating that marmoset vocalizations are not generated by
435   a fixed pattern generator but are regulated by a more versatile neural network.

436
437   **Chen, J. et al, 2021:** The authors report a specific population of estrogen receptor 1 (esr1)
438   positive neurons in the lateral preoptic area of the hypothalamus that can scale the amplitude of

439   vocalizations through a di-synaptic pathway acting via the periaqueductal grey. This neural circuit
440   is important for generating variability in USVs seen in social contexts.

441   **Conflict of Interest Statement:**
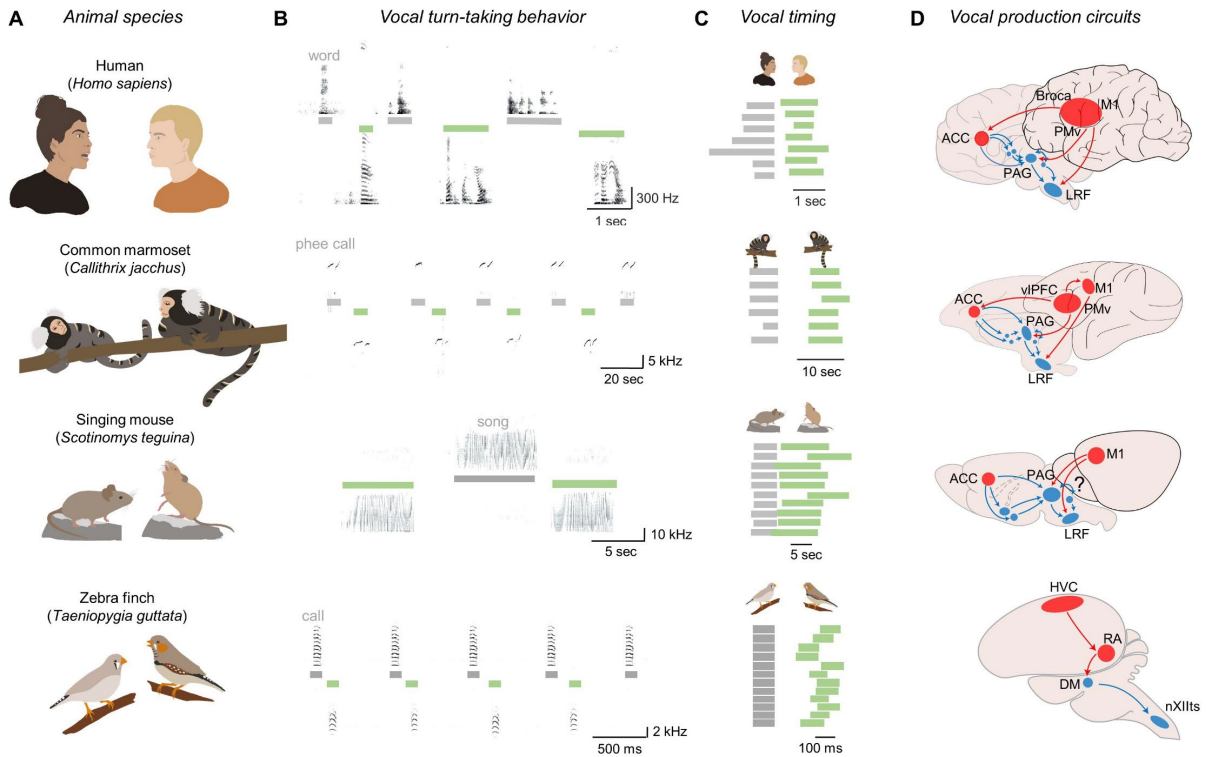442   The authors declare no conflict of interests.
443

450
451

Figures:

Figure 1 | Vocal turn-taking behavior of animals. A) Display of pairs of different animals vocally interacting. **B)** Sonograms of vocalizations used by the animals shown in **A**. Although the spectral features of the vocalizations have a species-specific signature, all animals take turns while interacting. **C)** Temporal profile of vocal turn taking. Vocalizations of animal 2 (green) aligned to the offset of the vocalizations of animal 1 (grey). Note the different times scales of these vocal exchanges. **D)** Neural circuits controlling vocal outputs: Broca: Broca's area; LRF: Laryngeal reticular formation; PMv: Premotor cortex (ventral); vPFC: ventral Pre-frontal cortex; PAG: Periaqueductal grey; M1: Primary motor cortex; ACC: Anterior cingulate cortex; nXIIts: Hyperglossal nucleus; DM: Dorsomedial nucleus of the intercollicular complex; RA: Robust nucleus of the arcopallium; HVC: proper name; Human data credit: Giacomo Costalunga & Neetash Mysuru, Marmoset data credit: Steffen Hage, Singing mice data credit: Clifford Harpole, Zebra finch data credit: Jonathan Benichov.
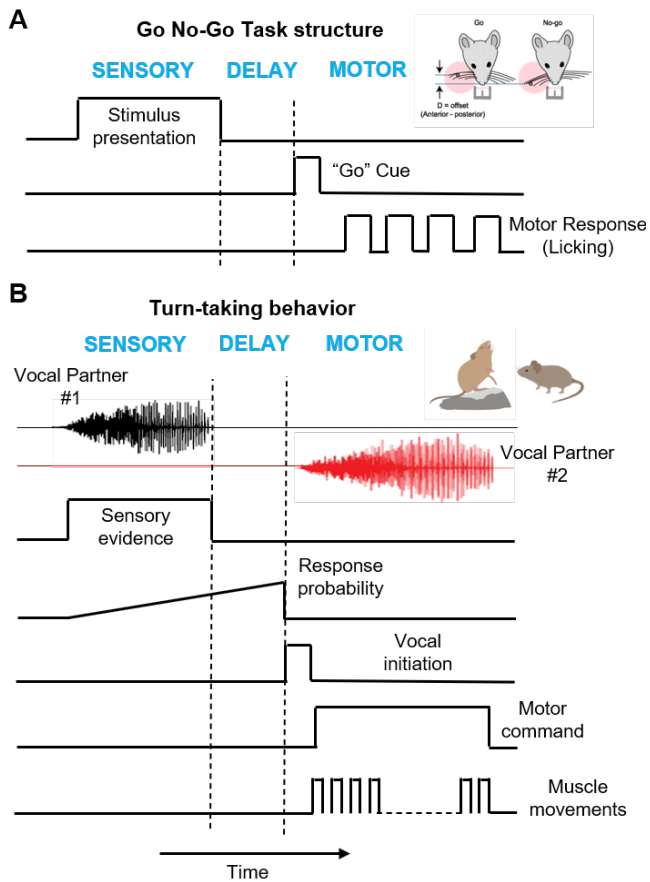
Figure 2 | Vocal turn-taking as paradigms for studying ethologically relevant sensorimotor transformations. A) Typical design of a Go No-Go task wherein a subject is presented with sensory stimuli followed by a delay period leading up to the motor response. Cartoon adapted from O'Conner *et. al.* 2010 **B)** Computations and their neural signatures underlying sensorimotor transformation during vocal turn-taking – from sensory evidence accumulation to decision making and finally leading up to motor commands and muscle movements. This example highlights vocal turn-taking behavior in the singing mice – a novel mammalian model system to study neural circuits for vocal communication. The minimal temporal overlap between the songs of the two mice conveniently delineates the behavior into sensory, delay and motor epochs, a feature typically engineered in many tasks.

492

493
494

References:

1. Catchpole, C.K. and P.J.B. Slater, *Bird song: Biological themes and variations*. 2008: Cambridge: Cambridge University Press.
2. De Gregorio, C., et al., *Categorical rhythms in a singing primate.* Current Biology, 2021. **31(20)**: p. R1379-R1380.
3. Konopka, G. and T.F. Roberts, *Animal Models of Speech and Vocal Communication Deficits Associated With Psychiatric Disorders.* Biol Psychiatry, 2016. **79**(1): p. 53-61.
4. Levinson, S.C., *Turn-taking in Human Communication--Origins and Implications for Language Processing.* Trends Cogn Sci, 2016. **20**(1): p. 6-14.
5. Levinson, S.C. and F. Torreira, *Timing in turn-taking and its implications for processing models of language.* Front Psychol, 2015. **6**: p. 731.
6. Stivers, T., et al., *Universals and cultural variation in turn-taking in conversation.* Proc Natl Acad Sci U S A, 2009. **106**(26): p. 10587-92.
7. Castellucci, G.A., et al., *A speech planning network for interactive language use.* Nature, 2022. **602**(7895): p. 117-122.
8. Donnelly, S. and E. Kidd, *The Longitudinal Relationship Between Conversational Turn-Taking and Vocabulary Growth in Early Language Development.* Child Dev, 2021. **92**(2): p. 609-625.
9. Bezerra, B.M. and A. Souto, *Structure and usage of the vocal repertoire of callithrix jacchus.* Int J Primatol., 2008. **29**: p. 671–701.
10. Takahashi, D.Y., D.Z. Narayanan, and A.A. Ghazanfar, *Coupled oscillator dynamics of vocal turn-taking in monkeys.* Curr Biol, 2013. **23**(21): p. 2162-8.
11. Dohmen, D. and S.R. Hage, *Limited capabilities for condition-dependent modulation of vocal turn-taking behavior in marmoset monkeys.* Behav Neurosci, 2019. **133**(3): p. 320-328.
12. Demartsev, V., et al., *Vocal Turn-Taking in Meerkat Group Calling Sessions.* Curr Biol, 2018. **28**(22): p. 3661-3666 e3.
13. Okobi, D.E., Jr., et al., *Motor cortical control of vocal interaction in neotropical singing mice.* Science, 2019. **363**(6430): p. 983-988.
14. Beckers, G.J. and M. Gahr, *Neural processing of short-term recurrence in songbird vocal communication.* PLoS One, 2010. **5**(6): p. e11129.
15. Hultsch, H. and D. Todt, *Repertoire Sharing and Song-Post Distance in Nightingales (Luscinia-Megarhynchos B).* Behavioral Ecology and Sociobiology, 1981. **8**(3): p. 183-188.
16. Hall, M.L., *A Review of Vocal Duetting in Birds.* Advances in the Study of Behavior, Vol 40, 2009. **40**: p. 67-121.
17. Fortune, E.S., et al., *Neural Mechanisms for the Coordination of Duet Singing in Wrens.* Science, 2011. **334**(6056): p. 666-670.
18. Coleman, M. and E. Fortune, *Duet singing in plain-tailed wrens.* Current Biology, 2018. **28**(11): p. R643-R645.
19. Hoffmann, S., et al., *Duets recorded in the wild reveal that interindividually coordinated motor control enables cooperative behavior.* Nature Communications, 2019. **10**.
20. Lemazina, A., et al., *The multifaceted vocal duets of white-browed sparrow weavers are based on complex duetting rules.* Journal of Avian Biology, 2021. **52**(9).

539   21.   Rivera-Caceres, K.D., et al., *Early development of vocal interaction rules in a duetting*
540   *songbird.* Royal Society Open Science, 2018. **5**(2).
541   22.   Benichov, J.I. and D. Vallentin, *Inhibition within a premotor circuit controls the timing of*
542   *vocal turn-taking in zebra finches.* Nat Commun, 2020. **11**(1): p. 221.
543   23.   Carter, G.G., et al., *Antiphonal calling allows individual discrimination in white-winged*
544   *vampire bats.* Animal Behaviour, 2008. **76**: p. 1343-1355.
545   24.   Knornschild, M., *Vocal production learning in bats.* Current Opinion in Neurobiology,
546   2014. **28**: p. 80-85.
547   25.   Fernandez, A.A., et al., *Babbling in a vocal learning bat resembles human infant*
548   *babbling.* Science, 2021. **373**(6557): p. 923-+.
549   26.   Knornschild, M., A.A. Fernandez, and M. Nagy, *Vocal information and the navigation of*
550   *social decisions in bats: Is social complexity linked to vocal complexity?* Functional
551   Ecology, 2020. **34**(2): p. 322-331.
552   27.   Nakahara, F. and N. Miyazaki, *Vocal exchanges of signature whistles in bottlenose*
553   *dolphins (Tursiops truncatus).* Journal of Ethology, 2011. **29**(2): p. 309-320.
554   28.   Drosopoulos, S. and M.F. Claridge, *Insect Sounds and Communication*. Vol. 1st Edition.
555   2005: Boca Raton CRC Press. 552.
556   29.   LaRue, K.M., et al., *Acoustic duetting in Drosophila virilis relies on the integration of*
557   *auditory and tactile signals.* Elife, 2015. **4**.
558   30.   Schulze, L., et al., *Transparent Danionella translucida as a genetically tractable*
559   *vertebrate brain model (vol 15, pg 977, 2018).* Nature Methods, 2018. **15**(12): p. 1126-
560   1126.
561   31.   Wilson, D.R., L.M. Ratcliffe, and D.J. Mennill, *Black-capped chickadees, Poecile*
562   *atricapillus, avoid song overlapping: evidence for the acoustic interference hypothesis.*
563   Animal Behaviour, 2016. **114**: p. 219-229.
564   32.   Hutfluss, A., et al., *Does song overlap signal aggressiveness? An experimental study with*
565   *repeated measures in free-ranging great tits.* Animal Behaviour, 2021. **179**: p. 199-211.
566   33.   Pika, S., et al., *Taking turns: bridging the gap between human and animal*
567   *communication.* Proc Biol Sci, 2018. **285**(1880).
568   34.   Wells, K.D., *Social-Behavior of Anuran Amphibians.* Animal Behaviour, 1977. **25**(Aug): p.
569   666-693.
570   35.   Fischer, J. and T. Price, *Meaning, intention, and inference in primate vocal*
571   *communication.* Neurosci Biobehav Rev, 2017. **82**: p. 22-31.
572   36.   Greenfield, M.D., M.K. Tourtellot, and W.A. Snedden, *Precedence effects and the*
573   *evolution of chorusing.* Proceedings of the Royal Society B-Biological Sciences, 1997.
574   **264**(1386): p. 1355-1361.
575   37.   Allen-Ankins, S. and L. Schwarzkopf, *Spectral overlap and temporal avoidance in a*
576   *tropical savannah frog community.* Animal Behaviour, 2021. **180**: p. 1-11.
577   38.   Evans, D.H. and S. Currie, *The physiology of fishes*. Vol. 5th edition. 2020: CRC Press.
578   39.   Bass, A.H. and B.P. Chagnaud, *Shared developmental and evolutionary origins for neural*
579   *basis of vocal-acoustic and pectoral-gestural signaling.* Proceedings of the National
580   Academy of Sciences of the United States of America, 2012. **109**: p. 10677-10684.
581   40.   Noordzij, J.P. and R.H. Ossoff, *Anatomy and physiology of the larynx.* Otolaryngologic
582   Clinics of North America, 2006. **39**(1): p. 1-+.

583   41.   Stein, R.C., *Modulation in Bird Sounds.* Auk, 1968. **85**(2): p. 229-&.

584   42.   Feng, N.Y. and A.H. Bass, *"Singing" Fish Rely on Circadian Rhythm and Melatonin for the*
585   *Timing of Nocturnal Courtship Vocalization.* Curr Biol, 2016. **26**(19): p. 2681-2689.

586   43.   Jarvis, E.D., *Evolution of vocal learning and spoken language.* Science, 2019. **366**(6461):
587   p. 50-54.

588   44.   Hage, S.R., *The role of auditory feedback on vocal pattern generation in marmoset*
589   *monkeys.* Curr Opin Neurobiol, 2020. **60**: p. 92-98.

590   45.   Nieder, A. and R. Mooney, *The neurobiology of innate, volitional and learned*
591   *vocalizations in mammals and birds.* Philos Trans R Soc Lond B Biol Sci, 2020. **375**(1789):
592   p. 20190054.

593   46.   Colquitt, B.M., et al., *Cellular transcriptomics reveals evolutionary identities of songbird*
594   *vocal circuits.* Science, 2021. **371**(6530).

595   47.   Hage, S.R. and A. Nieder, *Dual Neural Network Model for the Evolution of Speech and*
596   *Language.* Trends Neurosci, 2016. **39**(12): p. 813-829.

597   48.   Zhang, Y.S. and A.A. Ghazanfar, *A Hierarchy of Autonomous Systems for Vocal*
598   *Production.* Trends Neurosci, 2020. **43**(2): p. 115-126.

599   49.   Jurgens, U., *The role of the periaqueductal grey in vocal behaviour.* Behav Brain Res,
600   1994. **62**(2): p. 107-17.

601   50.   Tschida, K., et al., *A Specialized Neural Circuit Gates Social Vocalizations in the Mouse.*
602   Neuron, 2019. **103**(3): p. 459-472 e4.

603   51.   Jurgens, U., *The neural control of vocalization in mammals: a review.* J Voice, 2009.
604   **23**(1): p. 1-10.

605   52.   Michael, V., et al., *Circuit and synaptic organization of forebrain-to-midbrain pathways*
606   *that promote and suppress vocalization.* Elife, 2020. **9**.

607   53.   Chen, J., et al., *Flexible scaling and persistence of social vocal communication.* Nature,
608   2021. **593**(7857): p. 108-113.

609   54.   Coude, G., et al., *Neurons controlling voluntary vocalization in the macaque ventral*
610   *premotor cortex.* PLoS One, 2011. **6**(11): p. e26822.

611   55.   Hage, S.R., N. Gavrilov, and A. Nieder, *Cognitive control of distinct vocalizations in rhesus*
612   *monkeys.* J Cogn Neurosci, 2013. **25**(10): p. 1692-701.

613   56.   Takahashi, D.Y., A.R. Fenley, and A.A. Ghazanfar, *Early development of turn-taking with*
614   *parents shapes vocal acoustics in infant marmoset monkeys.* Philos Trans R Soc Lond B
615   Biol Sci, 2016. **371**(1693).

616   57.   Eliades, S.J. and C.T. Miller, *Marmoset vocal communication: Behavior and neurobiology.*
617   Dev Neurobiol, 2017. **77**(3): p. 286-299.

618   58.   Miller, C.T. and X. Wang, *Sensory-motor interactions modulate a primate vocal behavior:*
619   *antiphonal calling in common marmosets.* J Comp Physiol A Neuroethol Sens Neural
620   Behav Physiol, 2006. **192**(1): p. 27-38.

621   59.   Norcross, J.L. and J.D. Newman, *Context and gender-specific differences in the acoustic*
622   *structure of common marmoset (Callithrix jacchus) phee calls.* Am J Primatol, 1993.
623   **30**(1): p. 37-54.

624   60.   Miller, C.T., K. Mandel, and X. Wang, *The communicative content of the common*
625   *marmoset phee call during antiphonal calling.* Am J Primatol, 2010. **72**(11): p. 974-80.

626 61.   Miller, C.T. and A. Wren Thomas, *Individual recognition during bouts of antiphonal*
627   *calling in common marmosets.* J Comp Physiol A Neuroethol Sens Neural Behav Physiol,
628   2012. **198**(5): p. 337-46.
629 62.   Pomberger, T., et al., *Precise Motor Control Enables Rapid Flexibility in Vocal Behavior of*
630   *Marmoset Monkeys.* Curr Biol, 2018. **28**(5): p. 788-794 e3.
631 63.   Wang, X., et al., *Representation of a species-specific vocalization in the primary auditory*
632   *cortex of the common marmoset: temporal and spectral characteristics.* J Neurophysiol,
633   1995. **74**(6): p. 2685-706.
634 64.   Eliades, S.J. and X. Wang, *Sensory-motor interaction in the primate auditory cortex*
635   *during self-initiated vocalizations.* J Neurophysiol, 2003. **89**(4): p. 2194-207.
636 65.   Wolpert, D.M. and J.R. Flanagan, *Motor prediction.* Curr Biol, 2001. **11**(18): p. R729-32.
637 66.   Eliades, S.J. and X. Wang, *Neural substrates of vocalization feedback monitoring in*
638   *primate auditory cortex.* Nature, 2008. **453**(7198): p. 1102-6.
639 67.   Eliades, S.J. and J. Tsunada, *Auditory cortical activity drives feedback-dependent vocal*
640   *control in marmosets.* Nat Commun, 2018. **9**(1): p. 2540.
641 68.   Miller, C.T., et al., *Vocalization Induced CFos Expression in Marmoset Cortex.* Front Integr
642   Neurosci, 2010. **4**: p. 128.
643 69.   Miller, C.T., et al., *Responses of primate frontal cortex neurons during natural vocal*
644   *communication.* J Neurophysiol, 2015. **114**(2): p. 1158-71.
645 70.   Banerjee, A., S.M. Phelps, and M.A. Long, *Singing mice.* Curr Biol, 2019. **29**(6): p. R190-
646   R191.
647 71.   Nottebohm, F., T.M. Stokes, and C.M. Leonard, *Central control of song in the canary,*
648   *Serinus canarius.* J Comp Neurol, 1976. **165**(4): p. 457-86.
649 72.   Pfenning, A.R., et al., *Convergent transcriptional specializations in the brains of humans*
650   *and song-learning birds.* Science, 2014. **346**(6215): p. 1256846.
651 73.   Long, M.A., D.Z. Jin, and M.S. Fee, *Support for a synaptic chain model of neuronal*
652   *sequence generation.* Nature, 2010. **468**(7322): p. 394-9.
653 74.   Benichov, J.I., et al., *The Forebrain Song System Mediates Predictive Call Timing in*
654   *Female and Male Zebra Finches.* Curr Biol, 2016. **26**(3): p. 309-18.
655 75.   Coleman, M.J., et al., *Neurophysiological coordination of duet singing.* Proceedings of
656   the National Academy of Sciences of the United States of America, 2021. **118**(23).
657 76.   Krakauer, J.W., et al., *Neuroscience Needs Behavior: Correcting a Reductionist Bias.*
658   Neuron, 2017. **93**(3): p. 480-490.
659 77.   Jourjine, N. and H.E. Hoekstra, *Expanding evolutionary neuroscience: insights from*
660   *comparing variation in behavior.* Neuron, 2021. **109**(7): p. 1084-1099.
661 78.   Cisek, P., *Resynthesizing behavior through phylogenetic refinement.* Atten Percept
662   Psychophys, 2019. **81**(7): p. 2265-2287.
663