Quantification of developmental birdsong learning from the subsyllabic scale to cultural evolution

Dina Lipkind[1] and Ofer Tchernichovski

Department of Psychology, Hunter College, City University of New York, New York, NY 10065

Quantitative analysis of behavior plays an important role in birdsong neuroethology, serving as a common denominator in studies spanning molecular to system-level investigation of sensory-motor conversion, developmental learning, and pattern generation in the brain. In this review, we describe the role of behavioral analysis in facilitating cross-level integration. Modern sound analysis approaches allow investigation of developmental song learning across multiple time scales. Combined with novel methods that allow experimental control of vocal changes, it is now possible to test hypotheses about mechanisms of vocal learning. Further, song analysis can be done at the population level across generations to track cultural evolution and multigenerational behavioral processes. Complementing the investigation of song development with non-invasive brain imaging technology makes it now possible to study behavioral dynamics at multiple levels side by side with developmental changes in brain connectivity and in auditory responses.

animal communication | vocal development

Birdsong neuroethology is a small but diverse and influential subfield of neuroscience, studied in ∼100 laboratories, and covering a range of research areas spanning the molecular to the organismal levels of investigation. The quantification of singing behavior has been a common denominator across most of those studies. Over the past four decades, birdsong neuroethology has made tremendous gains by incorporating existing tools for representing and analyzing sound (1). The sonogram (Fig. 1A) (2–4) provides an excellent descriptive model of singing behavior, and over the past decade several analysis approaches were developed that allow automatic segmentation categorization and comparison of vocal sounds (5–9). In this review, we describe the role of behavioral analysis in the progress made in birdsong neuroethology research.

The past 20 y or so have seen rapid progress in understanding the neural and molecular processes involved in song. However, this progress was preceded by a long tradition of quantifying singing behavior (10, 11). Birdsong is uniquely amenable and attractive for behavioral analysis for a number of reasons: adult song is a structured behavior, repeated with a high degree of stereotypy (Fig. 1A). This makes it relatively easy to detect and characterize song structure and components (e.g., notes, syllables, motifs, and phrases) to align song renditions to each other, to assess the degree of similarity between songs, and, ultimately, to relate singing behavior to neuronal events and thus elucidate underlying neural mechanisms. We start by reviewing the role of behavioral quantification in achieving integration across levels of investigation. We then present a short outline of sound analysis techniques currently used for the tracking of developmental song learning, and an experimental approach for hypothesis-driven investigation of song development. Finally, we present a generalization of behavioral analysis to the population and cross-generation levels, and propose some ideas about new directions enabled by combining behavioral and brain imaging technologies.

Role of Behavioral Quantification in Achieving Integration Across Levels

The role of behavioral analysis in facilitating multiple levels of investigation can be demonstrated by presenting a few examples.

The genome of a songbird was sequenced earlier this year (12), motivated in part by the goal of discovering the genetic changes (as revealed by comparison with the chicken genome) that led to the evolution of vocal learning. In contrast to auditory learning (the ability to discriminate between sounds), vocal learning (the ability to imitate complex sounds) is rare in nature, but in birds it probably evolved in three out of the 25–28 known orders: songbirds, parrots, and hummingbirds (13–17). Aside from humans, vocal learning in mammals has been conclusively demonstrated only in dolphins, whales (18), and bats (19), but not in apes. Vocal learners share similar forebrain structures necessary to produce and acquire their learned vocalization (whether song or speech). In songbirds, these pathways (the song system) consist of a posterior song pathway necessary for producing learned song and an anterior song pathway mainly used for song acquisition (Fig. 2A). The auditory pathway (Fig. 2B) is connected to the vocal pathway and is directly involved in vocal learning (Fig. 2B) (13–17). Given the sparseness of vocal learning across taxa, it is remarkable that the same specific gene, FOXP2, associated with vocal learning is found in both humans and songbirds. A mutation in FOXP2 is associated with speech impairments in humans (20). FOXP2 is expressed in the songbird brain specifically in song system nuclei (21, 22). Interfering with its expression in the song system during song development causes subtle deficits in song learning, which are similar to those described in humans with FOXP2 mutation (23). The ability to quantitatively assess the accuracy of song imitation was crucial for making this discovery.

The attention of birdsong scientists to subtle yet biologically important modulation of singing behavior has led to several discoveries. For example, the zebra finch male (the "mouse" of birdsong neuroethology) learns and produces only a single song over its entire life. However, when courting a female, there are subtle differences in the performance of that song: singing becomes slightly faster and more stereotyped (24, 25). Female-directed vs. undirected song appear to differ also in their patterns of song system immediate-early gene expression. Undirected song is associated with a very strong expression of the immediate-early gene *egr-1* (also called ZENK) in the largest song nucleus, area X (Fig. 2A) (26), as well as down-regulation in the expression of FOXP2 (27). These changes in gene expression were not found during female-directed song. In addition, neuronal activity in area X was found to be higher and less stereotyped during undirected than directed song (28). The reason for these differences is most likely related to the role of area X in song learning. Area X belongs to the anterior pathway (Fig. 2A), a corticobasal ganglia circuitry that generates vocal exploration patterns that guide song learning and probably participate in song tuning and maintenance as well. The discovery of *egr-1* expression during undirected song was an early indicator of the role of the anterior
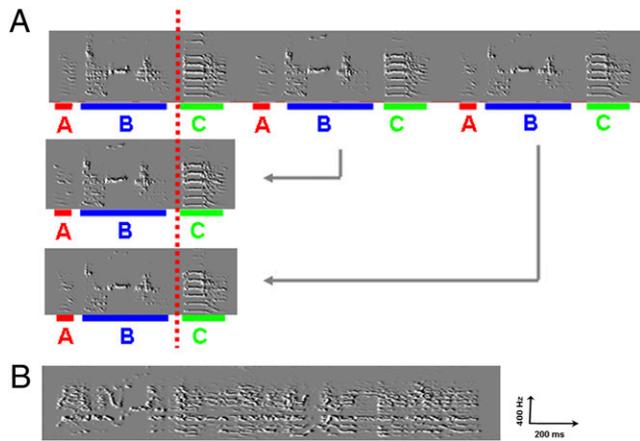
Fig. 1. Adult and juvenile song. (*A*) Spectral derivatives (4) provide a high-definition alternative to the classic sound spectrogram of a song of an adult (>100-d-old) zebra finch male. The song has a complex yet highly stereotyped structure of short sounds (syllables, designated by letters) arranged in repeating sequences (motifs). The song motifs are aligned below each other, showing the high stereotypy of their performance. Red dashed line indicates point of alignment. (*B*) Sound spectrogram of the vocalization (subsong) of a juvenile (40-d-old) zebra finch male. The sounds are complex, but not stereotyped. Vocal state keeps changing, but no obvious units, such as syllable types or motifs, can be detected.

pathway in the subtle exploratory noise (variability in song structure) that the bird generates during undirected song, but not during female-directed song, when perfect performance is desired (24, 25, 29). Here again, the ability to describe singing behavior in detail, and thus characterize the subtle differences between directed and undirected song, was the foundation for the ability to associate these differences with genetic and neural mechanisms.

Moving from the molecular to the cellular level, one of the most remarkable discoveries made in the 1980s in birdsong neuroethology was that of adult neurogenesis in the song system (30). Until then it was believed that there is no replacement of neurons in the adult brain. However, while investigating the neuroanatomy of the canary brain, seasonal changes in the size of song nuclei were observed, and were found to be strongly associated with seasonal changes in singing behavior. Canaries

are open-ended learners; they sing mainly in the fall and in the spring, and before each reproduction season they redevelop their songs, adding newly learned songs to their singing repertoire. Interestingly, large numbers of neurons are replaced by new neurons during the process of song learning. Those new neurons differentiate from radial glia and migrate into the song system to replace specific types of neurons, a process that has been hypothesized (although not yet fully corroborated) to facilitate song learning (31, 32). After adult neurogenesis was discovered in the songbird brain, numerous studies discovered it also in mammals [reviewed in Doetsch and Hen (33)] and in association with other behaviors—for example, spatial and social learning (34–36)—with high specificity in location and the neuronal types being replaced.

Another significant example is from electrophysiology studies, where there has been much recent progress in developing techniques that allow recording from single and multiple units in song nuclei of freely singing animals (37–42). Compared with researchers working on other learned behaviors, birdsong electrophysiologists have a major advantage because of the high stereotypy and accuracy of the complex song patterns of adult birds. As shown in Fig. 1*A*, it is easy to align song motifs and observe a replicable "singing state" at millisecond accuracy across renditions. This approach made it possible to associate spiking activity with momentary singing state and led to the discovery of an ultrasparse coding of song time (43), where each premotor neuron produces a single spike-train that accurately encodes a specific singing state, once per song motif, in a clockwork manner.

These examples illustrate the strength of birdsong neuroethology in combining detailed behavioral analysis with various levels of neuroscience to achieve cross-level understanding of the studied system. We next focus on song development, where classical behavior quantification methods have, to a large extent, failed to capture the behavioral process.

Difficulties with Quantification of Song Development

Although much progress was made between 1970 and 2000 in understanding mechanisms of song production and identifying neural correlates of song learning, it gradually became evident that the major gap in our understanding of the song system is at the level of song development. Just as studying song patterns gained from the availability of techniques for quantifying song structure, studying song development suffered from the lack of tools to capture the behavioral process of song learning. The problem was not merely technical. It is linked to limitations in methods and concepts routed deep in the field of ethology. The

Fig. 2. Vocal and auditory pathways of the songbird brain. (*A*) The vocal pathway, which consists of a posterior pathway necessary for producing learned song (brain nuclei composing it are shown in dark gray, and connections between them denoted with black arrows); and an anterior pathway necessary for song learning (brain nuclei composing it are shown in white, and connections between them denoted with white arrows). Dashed lines indicate connections between the two pathways. (*B*) The auditory pathway. Av, avalanche; CLM, caudal lateral mesopallium; CMM, caudal medial mesopallium; CN, cochlear nucleus; CSt, caudal striatum; DM, dorsal medial nucleus; DLM, dorsal lateral nucleus of the medial thalamus; E, entopallium; B, basorostralis; LLD, lateral lemniscus, dorsal nucleus; LLI, lateral lemniscus, intermediate nucleus; LLV, lateral lemniscus, ventral nucleus; MLd, dorsal lateral nucleus of the mesencephalon; LMAN, lateral magnocellular nucleus of the anterior nidopallium; area X, area X of the medial striatum; MO, oval nucleus of the mesopallium; NCM, caudal medial nidopallium; NIf, nucleus interface of the nidopallium; nXIIts, nucleus XII, tracheosyringeal part; Ov, ovoidalis; PAm, paraambiguus; RAm, retroambiguus; RA, robust nucleus of the arcopallium; SO, superior olive; Uva, nucleus uvaeformis. [Reproduced with permission from ref. 13 (Copyright 2004, The New York Academy of Sciences).]
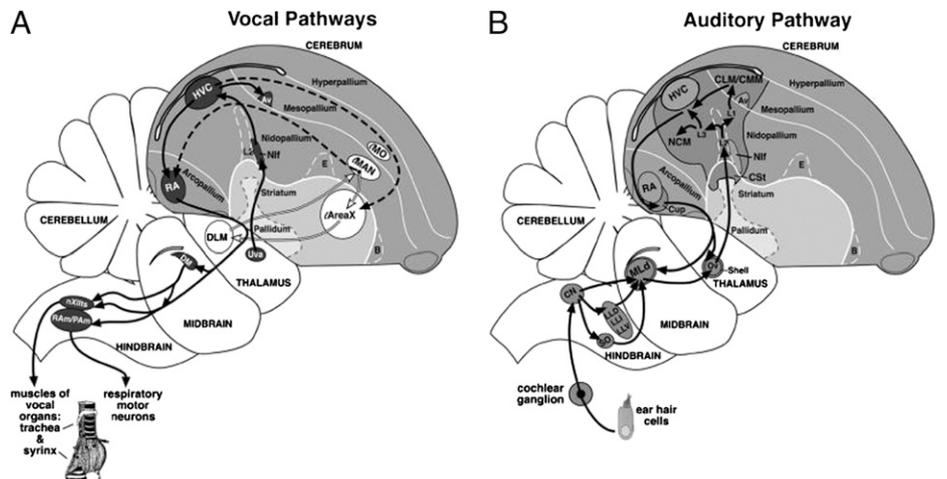
founders of birdsong neuroethology, Peter Marler and William H. Thorpe, were students of Nikolaas Tinbergen (44), and the behavioral analysis of song that formed its foundation was based on a combination of two approaches: first, obtaining a detailed descriptive model of singing behavior by applying sound spectral analysis to songs, namely the sonogram image (2); and second, analyzing the sonogram using the principles of classical ethology, which consist of segmenting spontaneous behavior into discrete units—fixed-action patterns. The idea that natural behavior has a fixed core is central to classical ethology (44, 45). Konrad Lorenz (45) saw himself as a comparative anatomist of behavior and sought to describe it as a sequence of its natural consisting units. Singing behavior fit this conceptual framework as few other behaviors did, because the song is often highly stereotyped and relatively easy to segment, and behavioral patterns are readily visible in the sonogram. However, the approach of identifying fixed behavioral patterns, although suitable for analyzing adult song, and for identifying stages in song development (e.g., detecting the pruning of syllables during late song development) (46), is considerably less effective in characterizing developmental song learning dynamically, as a continuous process.

The reason is that the structured and stereotyped adult song develops from highly variable and less structured juvenile vocalizations (Fig. 1*B*), which ethologists did not know how to analyze. In fact, much of the beauty of song development is in the process by which a graded signal, with a variety of states but no distinct structure, gradually transforms into a signal that includes discrete categories (syllables, motifs, etc.). The adult song can be treated analytically as a weakly symbolic signal, where one can identify distinct entities, count them, or look at their order, and so forth. None of these methods work with a graded signal, where the options for analysis, such as calculating continuous features and plotting their distributions, are completely different. Analyses

of graded and symbolic signals are effectively two different fields, and song development is a rare case where a natural transition from one to the other takes place and is readily accessible for analysis. The starting point of classical ethology has been to create a catalog of behavioral categories (ethogram). This approach is well-suited for symbolic signals. However, though adult song is easily described using distinct behavioral categories, this approach breaks down when attempting to describe the early graded signal and the process of its transformation into mature song.

Current Approaches to the Analysis of Song Development

Current approaches to the study of nonstereotyped behavior are strongly influenced by the rapid increase in computational power over the past decade. It is now easy to record and store the entire development of a song and explore the fine structure of millions of sounds produced during song learning. Therefore, instead of studying individual behavioral events, it became possible to study the distributions of continuous features over many sounds, and it became easier to investigate multiple time scales of song learning—from moment to moment (7, 47), over cycles of night sleep and morning singing (48), and over the entire learning trajectory (8). The analysis of the stereotyped adult song gained as well from analytic approaches based on large data sets of continuous features, revealing details of fine structure and variation that were not detectable via classical analysis (49, 50).

A conceptual framework was recently developed by Golani, Benjamini, and coworkers (51–53) at Tel Aviv University, which aims specifically at addressing the shortcomings of classical ethology while taking advantage of the current computational and informatics power. Their approach (which they used to study movement in rodents) is based on identifying and quantifying behavioral processes, not behavioral patterns. The idea is to use continuous series of low level features or metrics, which are then used to extract intrinsic units of behavior. Those "units" are not fixed (namely, they are not fixed patterns), but they nevertheless provide a natural framework for understanding the process being quantified. For example, when a mouse is exploring an arena, the intrinsic units are "exploratory excursions"—round trips from a home base and back. Each excursion is different, but analysis of the distribution of those excursions, as a continuous process, uncovers a global pattern of systematic incorporation of space segments, with escalation first around the walls and then toward the center (54). Although this analysis framework was developed for the investigation of movement patterns in mammals, it works quite well for the analysis of song development.

Fig. 3. Quantification of song development. (*A*) Features used for continuous description of song (7). Images on each side of the arrows demonstrate extreme values for each feature. (*B*) An example of an analyzed song rendition (spectrogram). The features shown in *A* are calculated for each millisecond of song (blue line, amplitude). Syllables (green) are defined by segmenting the amplitude time series with a certain threshold. Each syllable is characterized by its duration and its mean features values. The values for two syllables (in the red frames) are highlighted below. (*C*) The distribution of two of these mean values (entropy and FM) in the vocalizations of a juvenile (45 d old; *Left*) and of the same bird at a later developmental stage (90 d old; *Right*). The variable and unstructured vocalizations of the young bird are reflected in the broad features distribution, where no distinct clusters can be seen. The distribution at an older age reveals the emergence of clusters (syllable types).

Fig. 4. Altered-target training. Sound spectrograms of a bird's own song and of the source and target tutor songs (red frames). A juvenile male (43 d old; *Upper*) is trained with a source tutor song consisting of a single syllable type (AAAA). Once the bird has learned it (at the age of 55 d), training is switched to the target tutor song, in which a new syllable type is alternated with the old one (ABAB). This bird learned the target song on day 90 posthatch.

At the basis of our approach is an experimental setup that induces young birds to learn to sing under laboratory conditions, by keeping them in visual and acoustic isolation in sound attenuation chambers and exposing them to natural or synthetic song playbacks (7), which they imitate. The entire vocal output of the birds during the period of song development is recorded and stored, amounting to millions of syllables per bird. In the next stage, quantitative tools developed by Mitra and coworkers (55–57) at Bell Laboratories are applied to these vast datasets to fit descriptive models and track the emergence of the units of song structure. This process is demonstrated in Fig. 3. Low-level acoustic features, such as amplitude, pitch, frequency modulation (FM), and Wiener entropy (an estimate of the tonal purity of sounds; Fig. 3*A*) are calculated in sliding time windows of 10 ms of song, providing a near continuous description. The continuous sound is then segmented into syllables (segments of sound divided by silences). The mean values of the features within each syllable are then used to summarize its acoustic structure (syllable features; Fig. 3*B*). The distribution of syllable features is initially broad and unstructured (Fig. 3*C*, *Left*), but after a few days of training with song playbacks it becomes possible to see the emergence of clusters (Fig. 3*C*, *Right*). A typical zebra finch song consists of a two- to five-syllable–type vocabulary. Applying a clustering algorithm (57) to the multidimensional distribution of syllable features enables the identification of distinct clusters in feature space, corresponding to distinct syllable types. Usually, four to six acoustic features (e.g., syllable duration, mean and variance of pitch, FM, and Wiener entropy) are sufficient to successfully cluster all of the syllable types of a zebra finch song. The analysis of the distribution of syllable features hence reveals the process of emergence of discrete units of behavior—syllable

types—in the course of development (Fig. 3*C*). Once the song matures and becomes stereotyped, it can be described as a string of discrete syllable types, but during song development, one can investigate how clusters emerge, how they change shape, size, and location, and how the unclustered data are distributed.

There are some important differences between this description and that of identifying syllable types by visual inspection in small samples of songs. First, by looking at scatter plots it is easy to determine when each cluster (syllable category) appears and how it changes over development. Therefore, the behavioral units are obtained in an intrinsic and objective manner. Second, the categories remain transparent, so that beyond counting the number of renditions of each unit, or looking at the order at which they are performed, the experimenter can also characterize each rendition in detail, in reference to its cluster and in reference to the target. We can thus observe the dynamics of song development at multiple time scales. For example, the performance of song is affected by various factors, such as time of day (48, 49) and behavioral state [presence of female (24, 58) or hormonal levels (59, 60)]. These effects can be pinned down by quantifying the changes in the features and the variability of song syllables, which can be easily done by tracking changes in the location and size of their corresponding clusters in the feature distributions (Fig. 3*D*).

Notwithstanding the advantages of the approach detailed here, it is important to note that so far it has been applied mainly in studies of the zebra finch, the "mouse" of the birdsong neuroethology. However, just as the mouse is not a typical mammal, the zebra finch is not a typical songbird: in frequency structure its song is very complex, whereas in number of syllable types it is among the simplest. Other species of songbirds are better model

Fig. 5. The developmental trajectory of birds trained with altered-target training. (*A*) Sonograms of songs from different developmental days in two birds trained with AAAA→ABAB paradigms. The new syllable type (syllable B) appears first at one or both bout edges, and later in the center of the bout. Note that the two birds were trained with altered-target paradigms with a different syllable B—a harmonic stack in bird 1 and a broadband sound in bird 2. (*B*) Raster plots of bout edges for bird 1. Each line represents a song bout. Color indicates goodness of pitch at each millisecond of song. Bouts are arranged by temporal order from top to bottom; x axis, song time in ms; y axis, age in days. Bouts are aligned on the first rendition (*Left*) of syllable A and on its last rendition (*Right*). Alignment points are indicated by green arrows. The edges of song bouts are defined as the sounds before the first and after the last rendition of syllable A. Syllable B, which has high goodness of pitch, appears in red. It can be seen that the new syllable (syllable B) appears at the edges of bouts first and then at its center.

systems for studying different aspects of song behavior, such as the relation between production and perception or the structural complexity of songs. Swamp sparrows, which sing several song types, and countersing to rivals by matching song types, were recently used to demonstrate precise auditory-vocal mirroring in single cells of the song system nucleus HVC (Fig. 2*A*) (40). White-crowned sparrows have provided new insights into the relation between the complexity of learned songs and that of the auditory memories that guide learning (61, 62). Bengalese finches and canaries provide excellent model systems for studying higher levels of song organization (59–63). The sound analysis approach presented here is merely an example of implementing modern signal analysis techniques in neuroethology. There are now several alternative approaches to assess bioacoustic distances (64), including methods borrowed from speech analysis for the analysis of killer whale calls (65), as well as segmentation approaches for analysis of subsyllabic structure (49). Different approaches often complement each other, providing insight into different aspects of behavioral processes.

Hypothesis-Driven Song-Learning Research

Being able to observe vocal changes dynamically is useful but insufficient for understanding song development; to be able to test hypotheses about mechanism, one needs to gain some control over the vocal changes observed. At this front, two complementary approaches were recently developed. The first allows bypassing of the sensory-motor conversion (from song heard to song produced) to test directly how reinforcement signals can guide vocal changes. The second approach, developed by D.L. and presented here, is to give the birds imitation tasks designed to test specific hypotheses about song learning.

Guiding Vocal Changes by Real-Time Reinforcement Feedback. This technique was developed by Tumer and Brainard (66). It combines real-time sound analysis with specific manipulations of auditory feedback. Even the very stereotyped song has some variation in performance, so that the pitch of certain syllables has measurable jitter over renditions. Remarkably, a mild negative reinforcement, such as a burst of white noise, is sufficient to drive adaptive vocal changes. For example, a noise burst targeted at renditions of a specific syllable type only when the pitch happens to be a bit higher results in an almost immediate downshift in the pitch of the next renditions of that syllable type. Once the bird develops this avoidance response, the experimenter can shift the disruption target slightly more, and shape vocal changes over days in any desired direction solely by reinforcement. This effect indicates that the minor fluctuations in pitch across renditions are centrally driven and that the bird can associate them with the outcome of the song so as to make vocal changes that avoid disruptions. Such reinforcement feedback was recently used to investigate the role of the anterior forebrain pathway (Fig. 2*B*) in guiding vocal changes (67). It is easy to see how this technique can be used to test a variety of hypotheses about the saliency of song features and online error correction. The limitations of this approach are that it bypasses the sensory-motor conversion and does not include an explicit imitation task for the bird to perform.

Altered-Target Training. The fact that mature song develops from unstructured and variable vocalizations makes it hard to ask specific questions about the process. However, we can standardize both the starting point and endpoint of development by training young birds sequentially with two custom-designed song playbacks. A juvenile is first trained with one song (the source), and once he has learned it, training is switched to a second song (the target). The source and target songs can be designed to ask specific questions about song development. An example of this approach is presented in Fig. 4. We trained birds, starting on day 43 posthatch, with a source song consisting of a repeating syllable type (AAAA; Fig. 4). Once the source was learned (usually within 10 d), the playback was switched to the target song, which consists of two alternating syllable types: the source syllable and a new syllable (ABAB; Fig. 4). Thus, the bird is presented with a specific learning task that necessitates insertion of syllable type B between the renditions of A. By examining the developmental trajectory that leads from source to target, we can distinguish between alternative hypotheses about how the learning task is accomplished.

Visual inspection of sonograms can be used to obtain a first impression of a bird's developmental trajectory during the period of transition from source to target song. For example, Fig. 5*A* shows sonograms from different developmental days of two birds trained with the AAAA→ABAB task. Syllable B is absent early in

Fig. 6. Evolution of song culture. (*A*, *Upper*) Two examples of WT zebra finch songs. (*Lower*) Two examples of songs of birds reared in isolation (isolates). Colored lines indicate different syllable types within each bird's song. Despite variability among individuals, isolate songs differ from WT songs in syllable structure, duration, and order. (*B*) Experimental setup. An isolate tutor is kept singly with a pupil, who later serves as a tutor to the next pupil, and so on. (*C*, *Upper*) Example of a song of an isolate tutor containing an abnormally long syllable (red frame). (*Lower*) The same syllable type was shortened in the pupil's song. (*D*, *Upper*) Example of a song of another isolate tutor, containing back-to-back renditions of syllable types (marked by letters). (*Lower*) The pupil copied the syllable types but arranged them in a different, motif-like, order. Reproduced with permission from ref. 70.

Fig. 7. Quantification of convergence from isolate to WT song across generations. (*A*) Distribution of song features in WT songs and the songs' isolates and their pupils. Axes, the first two principal components (PC1 and PC2) derived from PCA of spectral features. The features distribution of WT songs is shown in purple. Shades of purple indicate the distance from the center of the WT cluster (darker is closer to center). Arrows of the same color indicate a certain learning dynasty (i.e., an isolate tutor and his pupils). (*Left*) Arrows connect each isolate tutor to its first generation pupils, showing pupils to be more WT-like. (*Right*) Arrows connect pupils from successive learning generations, showing a continued convergence toward WT-like song features. (*B*) Examples of a song syllable from an isolate tutor, and of successive generations of its pupils (from top to bottom). An abnormally long part of the syllable (red frames) is gradually shortened over generations. Reproduced with permission from ref. 70.

the transition process (day 55). At a later age (day 62), syllable B can be seen at one or both of the bout edges, and still later in development (day 70) it appears also in the bout center. However, sonograms do not permit a continuous visualization of this process. To do that, we plot song bouts in raster plots (Fig. 5*B*), which allow us to view the entire developmental trajectory at a glance. Furthermore, we can focus on desired parts of the song by an appropriate choice of alignment method and the spectral feature to be plotted. For example, to look specifically at the bout edges, we align the bout rasters on the first rendition of syllable A (which

allows us to get a clear view of the start of the bout; Fig. 5*B*, *Left*) and on its last rendition (which shows the end of the bout; Fig. 5*B*, *Right*). To detect the location within the song bout of syllable B (in this case a harmonic stack), the color code represents the goodness of pitch, a song feature that detects harmonic structure. It is therefore easy to track syllable B's position in the song bout and how it changes during development, because it appears in red in the raster plots. In the presented example, it can be seen that the new syllable appears first at the edges (start and end) of the bout, and later in development, also in the middle of the bout.

The advantage of altered target training is that it increases our control over the birds' complex developmental trajectories by standardizing the start point, as well as the end point, of song development. This method can be used to test hypotheses about various aspects of song development, by designing source and target songs that require birds to perform different kinds of learning tasks (e.g., changing syllable syntax, phonology, song rhythm, etc.).

Song Development over Generations: Evolution of Song Culture

In the individual bird, song learning occurs over multiple time scales: from moment to moment during singing and probably during other social interactions as well; during night sleep (48); and during developmental stages. However, song may develop over even longer time scales at the group level. Songbirds learn their songs by imitating an adult tutor. Imitation increases song similarity within a group, but there are also divergent forces (68) that drive considerable variability in song among individuals; moreover, there are regional dialects among groups of birds from the same species (69). Thus, birds have a song culture that develops over generations. How does song culture come about and what are the factors that influence it? Recently, experimental manipulations and sound analysis techniques were generalized to study the development of song culture in the laboratory (70). A key factor in being able to do this is the phenomenon of isolate songs. Young male birds that are reared in isolation (isolates) develop songs that differ markedly from the songs of normally reared birds (Fig. 6*A*) (71). Using an isolate bird as a tutor, and then using his pupil as a tutor, and so on recursively (Fig. 6*B*), it is possible to track a multigenerational trajectory of vocal changes. At the end of the experiment, we have recordings of the songs learned by males from successive generations—the isolate founder, his pupil, the pupil's pupil, and so on. Strikingly, within three to four learning generations, the songs evolve from isolate-like songs to wild-type (WT) songs without any external input. We can see this effect both at the syllable level, where abnormally long and unstructured isolate syllables evolve into shorter and more structured sounds (Fig. 6*C*), and at the level of the entire song, where back-to-back renditions of the same syllable—characteristic of some isolate songs—is replaced by a motif-like syntax of different syllable
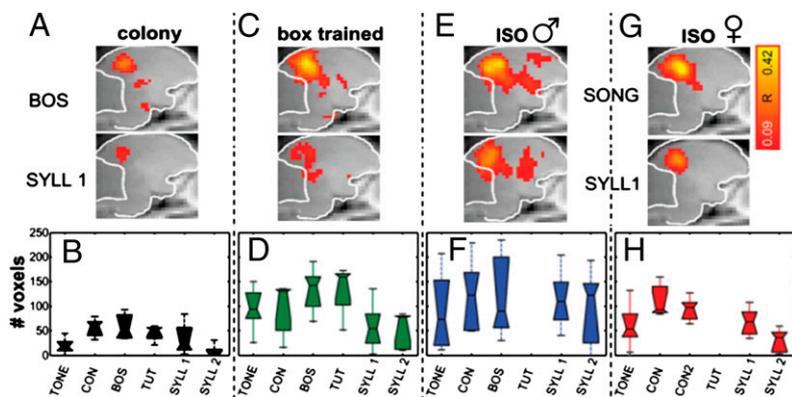
Fig. 8. Differences in BOLD activation to different stimuli. (*A*) BOLD activation maps for bird's own song (BOS) and repeated syllable (SYLL1) in colony males show stronger activation to BOS. Each map shows average activity in left and right hemisphere 0.5 mm from midline; color scale shows correlation coefficient. (*B*) Box plot summarizing BOLD activation across stimuli in colony males. Area of activation varies significantly across stimuli with greater activation to songs (ANOVA, $P < 0.01$). Stimuli: 2 kHz pure tone, conspecific song (CON), BOS, tutor song (TUT), and song syllables (SYLL 1, SYLL 2). All conspecific songs were produced by unfamiliar, colony-raised birds. (*C* and *D*) Same as *A* and *B* for box-trained males ($P < 0.05$). (*E* and *F*) Same as *A* and *B* for isolated males ($P = 0.9$; the TUT stimulus is absent in *F*, because isolates were not exposed to a tutor). (*G* and *H*) Same as *A* and *B* for iso-late females ($P < 0.05$) with an additional conspecific song (CON2) to balance cross-group comparisons. Reproduced from ref. 76 with permission from John Wiley & Sons, Inc. (Copyright 2010).

Lipkind and Tchernichovski

types performed in a fixed serial order (Fig. 6D). Because this happens without any external influence, the conclusion is that this process is a result of innate (genetically determined) imitation biases that gradually accumulate over learning generations. Isolate and WT songs differ along several dimensions, such as syllable phonology, duration of acoustic states (longer in isolates; Fig. 6A), and song syntax (isolates tend to repeat syllable types, whereas WT birds tend to alternate them). Performing principal components analysis (PCA) on cumulative distributions of song features shows a separation between the isolate and WT songs. Annotating the WT song distribution in purple shading and plotting vectors from each isolate tutor to its pupil (Fig. 7A, Left) show clearly that the arrows are leading toward WT distribution. Plotting multigenerational vectors (head to tail; Fig. 7A, Right) shows that across generations, the song features penetrate more deeply into the WT distribution. In this way it is possible to quantify the multigenerational development of song culture. This quantification method confirms the impression received from examining song renditions of pupils from successive learning generations (Fig. 7B).

Combining Behavioral Studies of Song Development with Neural Interventions and Brain Imaging

In the previous sections we have reviewed a set of experimental and analysis tools for the behavioral study of song development across relevant time scales. These tools can be used to relate behavioral to neural processes. Two approaches are especially useful for this: first, it is now possible to temporarily inactivate specific song nuclei so as to compare the distribution of features in the developing song when certain parts of the song system are turned on and off (72). Second, NMR-based brain imaging techniques such as fMRI and diffusion tensor imaging (DTI) are especially appropriate for looking at developmental changes not only at the motor side but also at the sensory and perceptual side. The fMRI blood oxygenation level-dependent (BOLD) responses to auditory stimulus correlate mostly with the local field potential (73), which does not measure the output of the neural computation, but is indicative of the synaptic and dendritic processes. Therefore, BOLD responses may capture the dynamics of computation when the bird perceives songs and interprets them, whereas spikes often capture the output of this process. Unlike many other methods, NMR imaging can be used to perform multiple testing on the same individual over long time periods, and thus capture the neural processes associated with the developing song. In this review we focus on NMR imaging that has the advantage of obtaining a global picture of brain activity and structure at different stages of song development, which is not possible with more invasive methods, such as single or multicell recordings. In addition, we can study how auditory responses are influenced by prior song production, to reveal how developing perception and production mechanisms influence each other. Finally, brain imaging and behavioral techniques can be combined to study environmental influences (e.g., social interactions) on song development and on the development of auditory perception. The major weakness of these techniques, besides their coarse spatial and temporal resolution, is that they cannot be used during singing behavior, and are limited to studies of auditory responses and brain connectivity in anesthetized or sedated animals.

fMRI and DTI are complementary: the first captures changes in blood flow, which is correlated with brain activation, and the second captures changes in white matter density, which indicate changes in brain connectivity. The current NMR technology with strong magnets (7T and 9T) now makes it possible to obtain spatial resolutions that are accurate enough to study a brain as small as that of the songbird (74). Even 3T magnets are sufficient to detect how early social and song-learning experience affects patterns of song-specific auditory responses (75). An example of this approach is a recent study that used brain imaging to test how vocal and auditory experience and social environment shape auditory responses (76). The auditory responses of isolate males versus males exposed to song during development were measured using fMRI and event-related potentials. Interestingly, auditory responses were shaped by song learning: males reared in a seminatural colony showed stimulus-specific auditory responses, responding more to their own and their tutor's song, less to a pure tone, and even less to calls (Fig. 8 A and B). Similar results were obtained for socially isolated males exposed solely to song playbacks, sufficient to induce song imitation. In contrast, isolate males that were not given an opportunity to imitate a song (did not hear song playbacks) did not show significant response selectivity across stimuli (Fig. 8C). Female zebra finches do not sing, but we found that isolate females that did not hear song playbacks nevertheless developed stimulus-specific responses similar to those of normally reared males (Fig. 8D), indicating that, in females, there is a stronger hard-wired specificity of auditory responses compared with males.

Future studies combining imaging and behavioral techniques promise to elucidate the role of social, auditory, and vocal experience in the development of song and of auditory perception. The incorporation of the DTI technology into these studies may even allow tracking anatomical changes in brain connectivity as well brain activation during song development. A particularly fascinating direction of this research is exploring the relation between changes in brain function and anatomy and distinct events in song development, such as the emergence of syllable types. By controlling the social environment, we can study the influence of a gradient of social complexity (from isolates, to male-male or male-female conspecific pairs, to larger social groups) on vocal and auditory development. A similar approach can be used for cross-generational studies at the cultural evolution level.

In addition to NMR technology, various brain imaging technologies at microspatial scales, such as two-photon imaging (77), and mapping of gene expression dynamics across brain regions (78, 79) are rapidly evolving. As with fMRI, those techniques are incorporated at different scales into birdsong neuroethology studies to capture different aspects of neural computation, output, and structural plasticity. The songbird offers a model system for combining these methods with analysis of behavior across multiple time scales.

Conclusions

In this review we aimed to outline the role of behavioral analysis in birdsong neuroethology, its history, present state, and future directions. Throughout its history, the success of the field strongly depended on the availability of a detailed and reliable description of behavior. Such description is achieved through a combination of several factors: experimental setups that display the behavioral process in a clear and replicable manner; data acquisition technologies that allow the collection and storage of large data sets; and the use of quantitative tools that provide a detailed description and categorization of the behavior. Major drawbacks in the past were the lack of an appropriate descriptive model of the process of song development, and of experimental setups amenable to testing hypotheses about the model or process. These drawbacks are now being addressed by a variety of experimental and quantification tools, making it possible to describe song development on multiple time scales. Using series of near-continuous song features, rather than a priori categories (ethograms), for data categorization and analysis allows for improved descriptions of the process of song development. Combining these sound analysis techniques with brain imaging promises to provide a global picture of the changes in brain activation and connectivity that bring song development about.

1. Percival DB, Walden AT (1993) *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques* (Cambridge Univ Press, Cambridge, U.K.).
2. Thorpe WH (1954) The process of song-learning in the Chaffinch as studied by means of the sound spectrograph. *Nature* 173:465–469.
3. Thomson DJ (1982) Spectrum estimation and harmonic analysis. *Proc IEEE* 70:1055–1096.
4. Thomson DJ (1990) Quadratic-inverse spectrum estimates: Applications to paleoclimatology. *Philos Trans R Soc Lond A* 332:539–597.
5. Kogan JA, Margoliash D (1998) Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *J Acoust Soc Am* 103:2185–2196.
6. Du P, Troyer T (2006) A segmentation algorithm for zebra finch song at the note level. *Neurocomputing* 69:1375–1379.
7. Tchernichovski O, Lints TJ, Deregnaucourt S, Cimenser A, Mitra PP (2004) Studying the song development process: Rationale and methods. *Ann N Y Acad Sci* 1016:348–363.
8. Tchernichovski O, Mitra PP, Lints T, Nottebohm F (2001) Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science* 291:2564–2569.
9. 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.
10. Hinde RA, ed (1969) *Bird Vocalizations: Their Relations to Current Problems in Biology and Psychology* (Cambridge Univ Press, Cambridge, U.K.).
11. Kroodsma DE, Miller EH, eds (1982) *Acoustic Communication in Birds* (Academic, New York).
12. Warren WC, et al. (2010) The genome of a songbird. *Nature* 464:757–762.
13. Reiner A, Perkel DJ, Mello CV, Jarvis ED (2004) Songbirds and the revised avian brain nomenclature. *Ann N Y Acad Sci* 1016:77–108.
14. Jarvis ED (2007) Neural systems for vocal learning in birds and humans: A synopsis. *J Ornithol* 148:35–44.
15. Jarvis ED, et al. (2000) Behaviourally driven gene expression reveals song nuclei in hummingbird brain. *Nature* 406:628–632.
16. Matsunaga E, Kato M, Okanoya K (2008) Comparative analysis of gene expressions among avian brains: A molecular approach to the evolution of vocal learning. *Brain Res Bull* 75:474–479.
17. Jarvis ED, Mello CV (2000) Molecular mapping of brain areas involved in parrot vocal communication. *J Comp Neurol* 419:1–31.
18. Janik VM, Slater PJB (1997) Vocal learning in mammals. *Adv Study Behav* 26:59–99.
19. Knörnschild M, Nagy M, Metz M, Mayer F, Helversen OV (2009) Complex vocal imitation during ontogeny in a bat. *Biol Lett* 6:156–159.
20. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519–523.
21. Haesler S, et al. (2004) FoxP2 expression in avian vocal learners and non-learners. *J Neurosci* 24:3164–3175.
22. Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA (2004) Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *J Neurosci* 24:3152–3163.
23. Haesler S, et al. (2007) Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. *PLoS Biol* 5:e321.
24. Kao MH, Brainard MS (2006) Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J Neurophysiol* 96:1441–1455.
25. Sakata JT, Hampton CM, Brainard MS (2008) Social modulation of sequence and syllable variability in adult birdsong. *J Neurophysiol* 99:1700–1711.
26. Jarvis ED, Scharff C, Grossman MR, Ramos JA, Nottebohm F (1998) For whom the bird sings: Context-dependent gene expression. *Neuron* 21:775–788.
27. Teramitsu I, White SA (2006) FoxP2 regulation during undirected singing in adult songbirds. *J Neurosci* 26:7390–7394.
28. Hessler N a Doupe a J (1999) Social context modulates singing-related neural activity in the songbird forebrain. *Nat Neurosci* 2:209–211.
29. Hampton CM, Sakata JT, Brainard MS (2009) An avian basal ganglia-forebrain circuit contributes differentially to syllable versus sequence variability of adult Bengalese finch song. *J Neurophysiol* 101:3235–3245.
30. Goldman SA, Nottebohm F (1983) Neuronal production, migration, and differentiation in a vocal control nucleus of the adult female canary brain. *Proc Natl Acad Sci USA* 80:2390–2394.
31. Alvarez-Buylla A, Kirn JR (1997) Birth, migration, incorporation, and death of vocal control neurons in adult songbirds. *J Neurobiol* 33:585–601.
32. Wilbrecht L, Williams H, Gangadhar N, Nottebohm F (2006) High levels of new neuron addition persist when the sensitive period for song learning is experimentally prolonged. *J Neurosci* 26:9135–9141.
33. Doetsch F, Hen R (2005) Young and excitable: The function of new neurons in the adult mammalian brain. *Curr Opin Neurobiol* 15:121–128.
34. Barnea A, Nottebohm F (1994) Seasonal recruitment of hippocampal neurons in adult free-ranging black-capped chickadees. *Proc Natl Acad Sci USA* 91:11217–11221.
35. Epp JR, Haack AK, Galea LAM (2009) Task difficulty in the Morris water task influences the survival of new neurons in the dentate gyrus. *Hippocampus* 876:866–876.
36. Adar E, Nottebohm F, Barnea A (2008) The relationship between nature of social change, age, and position of new neurons and their survival in adult zebra finch brain. *J Neurosci* 28:5394–5400.
37. McCasland JS, Konishi M (1981) Interaction between auditory and motor activities in an avian song control nucleus. *Proc Natl Acad Sci USA* 78:7815–7819.
38. Chi Z, Margoliash D (2001) Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron* 32:899–910.
39. Dave AS, Margoliash D (2000) Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* 290:812–816.
40. Prather JF, Peters S, Nowicki S, Mooney R (2008) Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* 451:305–310.
41. Fee MS, Leonardo A (2001) Miniature motorized microdrive and commutator system for chronic neural recording in small animals. *J Neurosci Methods* 112:83–94.
42. Vu ET, Schmidt MF, Mazurek ME (1998) Interhemispheric coordination of premotor neural activity during singing in adult zebra finches. *J Neurosci* 18:9088–9098.
43. Hahnloser RHR, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419:65–70.
44. Lorenz K (1981) *The Foundations of Ethology* (Springer, New York).
45. Tinbergen N (1951) *The Study of Instinct* (Oxford Univ Press, New York).
46. Marler P (1991) Song-learning behavior: The interface with neuroethology. *Trends Neurosci* 14:199–206.
47. Derégnaucourt S, et al. (2004) Song development: In search of the error-signal. *Ann N Y Acad Sci* 1016:364–376.
48. Derégnaucourt S, Mitra PP, Fehér O, Pytte C, Tchernichovski O (2005) How sleep affects the developmental learning of bird song. *Nature* 433:710–716.
49. Glaze CM, Troyer TW (2006) Temporal structure in zebra finch song: Implications for motor coding. *J Neurosci* 26:991–1005.
50. Glaze CM, Troyer TW (2007) Behavioral measurements of a temporally precise motor code for birdsong. *J Neurosci* 27:7631–7639.
51. Drai D, Golani I (2001) SEE: A tool for the visualization and analysis of rodent exploratory behavior. *Neurosci Biobehav Rev* 25:409–426.
52. Benjamini Y, et al. (2010) Ten ways to improve the quality of descriptions of whole-animal movement. *Neurosci Biobehav Rev* 34:1351–1365.
53. Benjamini Y, Fonio E, Galili T, Havkin GZ, Golani I (2011) Quantifying the buildup in extent and complexity of free exploration in mice. *Proc Natl Acad Sci USA*, 10.1073/pnas.1014837108.
54. Fonio E, Benjamini Y, Golani I (2009) Freedom of movement and the stability of its unfolding in free exploration of mice. *Proc Natl Acad Sci USA* 106:21335–21340.
55. Ho C, Pesaran B, Fee M, Mitra P (1998) Characterization of the structure and variability of zebra finch song elements. *Proceedings of the Fifth Joint Symposium on Neural Computation* (Inst for Neural Computation, San Diego), pp 76–83.
56. Tchernichovski O, Mitra PP (2002) Towards quantification of vocal imitation in the zebra finch. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 188:867–878.
57. Tchernichovski O, Mitra PP (2004) *Sound Analysis Pro User Manual*. Available at http://ofer.sci.ccny.cuny.edu.
58. Kojima S, Doupe AJ (2010) Social performance reveals unexpected vocal competency in young songbirds. *Proc Natl Acad Sci USA*, 10.1073/pnas.1010502108.
59. Marler P, Peters S, Wingfield J (1987) Correlations between song acquisition, song production, and plasma levels of testosterone and estradiol in sparrows. *J Neurobiol* 6:531–548.
60. Arnold AP (1997) Experimental analysis of sexual differentiation of the zebra finch brain. *Brain Res Bull* 44:503–507.
61. Rose GJ, et al. (2004) Species-typical songs in white-crowned sparrows tutored with only phrase pairs. *Nature* 432:753–758.
62. Plamondon SL, Rose GJ, Goller F (2010) Roles of syntax information in directing song development in white-crowned sparrows (Zonotrichia leucophrys). *J Comp Psychol* 124:117–132.
63. Gardner TJ, Naef F, Nottebohm F (2005) Freedom and rules: The acquisition and reprogramming of a bird's learned song. *Science* 308:1046–1049.
64. Ranjard L, et al. (2010) Bioacoustic distances between the begging calls of brood parasites and their host species: A comparison of metrics and techniques. *Behav Ecol Sociobiol* 64:1915–1926.
65. Shapiro DA, Tyack PL, Seneff S (2010) Comparing call-based versus subunit-based methods for categorizing Norwegian killer whale, Orcinus orca, vocalizations. *Anim Behav* 81:377–386.
66. Tumer EC, Brainard MS (2007) Performance variability enables adaptive plasticity of "crystallized" adult birdsong. *Nature* 450:1240–1244.
67. Andalman AS, Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc Natl Acad Sci USA* 106:12518–12523.
68. Tchernichovski O, Nottebohm F (1998) Social inhibition of song imitation among sibling male zebra finches. *Proc Natl Acad Sci USA* 95:8951–8956.
69. Marler P, Tamura M (1964) Culturally transmitted patterns of vocal behavior in sparrows. *Science* 146:1483–1486.
70. Fehér O, Wang H, Saar S, Mitra PP, Tchernichovski O (2009) De novo establishment of wild-type song culture in the zebra finch. *Nature* 459:564–568.
71. Williams H, Kilander K, Sotanski ML (1993) Untutored song, reproductive success and song learning. *Anim Behav* 45:695–705.
72. Olveczky BP, Andalman AS, Fee MS (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol* 3:e153.
73. Berens P, Logothetis N, Tolias A (2010) Local field potentials, BOLD and spiking activity—relationships and physiological mechanisms. Nature Precedings. Available at http://hdl.handle.net/10101/npre.2010.5216.1.
74. Van der Linden A, Van Meir V, Boumans T, Poirier C, Balthazart J (2009) MRI in small brains displaying extensive plasticity. *Trends Neurosci* 32:257–266.
75. Voss HU, et al. (2007) Functional MRI of the zebra finch brain during song stimulation suggests a lateralized response topography. *Proc Natl Acad Sci USA* 104:10667–10672.
76. Maul KK, et al. (2010) The development of stimulus-specific auditory responses requires song exposure in male but not female zebra finches. *Dev Neurobiol* 70:28–40.
77. Roberts TF, Tschida KA, Klein ME, Mooney R (2010) Rapid spine stabilization and synaptic enhancement at the onset of behavioural learning. *Nature* 463:948–952.
78. Gobes SMH, Zandbergen MA, Bolhuis JJ (2010) Memory in the making: Localized brain activation related to song learning in young songbirds. *Proc Biol Sci* 277:3343–3351.
79. Horita H, Wada K, Rivas MV, Hara E, Jarvis ED (2010) The dusp1 immediate early gene is regulated by natural stimuli predominantly in sensory input neurons. *J Comp Neurol* 518:2873–2901.