

Auditory-induced neural dynamics in sensory-motor circuitry predict learned temporal and sequential statistics of birdsong

Kristofer E. Bouchard^{a,1} and Michael S. Brainard^{b,c,d}

^aBiological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; ^bDepartment of Physiology, University of California, San Francisco, CA 94158; ^cCenter for Integrative Neuroscience, University of California, San Francisco, CA 94158; and ^dHoward Hughes Medical Institute, University of California, San Francisco, CA 94158

Edited by Thomas D. Albright, The Salk Institute for Biological Studies, La Jolla, CA, and approved June 28, 2016 (received for review May 1, 2016)

Predicting future events is a critical computation for both perception and behavior. Despite the essential nature of this computation, there are few studies demonstrating neural activity that predicts specific events in learned, probabilistic sequences. Here, we test the hypotheses that the dynamics of internally generated neural activity are predictive of future events and are structured by the learned temporal-sequential statistics of those events. We recorded neural activity in Bengalese finch sensory-motor area HVC in response to playback of sequences from individuals' songs, and examined the neural activity that continued after stimulus offset. We found that the strength of response to a syllable in the sequence depended on the delay at which that syllable was played, with a maximal response when the delay matched the intersyllable gap normally present for that specific syllable during song production. Furthermore, poststimulus neural activity induced by sequence playback resembled the neural response to the next syllable in the sequence when that syllable was predictable, but not when the next syllable was uncertain. Our results demonstrate that the dynamics of internally generated HVC neural activity are predictive of the learned temporal-sequential structure of produced song and that the strength of this prediction is modulated by uncertainty.

sequences | prediction | uncertainty | electrophysiology | birdsong

Most complex behaviors, such as vocalizations of humans and birds, unfold over time as coordinated sequences (1–4). Furthermore, the sensory world exhibits sequential structure, and so perception also reflects processing a series of sequential events (5–8). The statistical regularities of naturally occurring sequences potentially enable prediction of upcoming sensory/behavioral events based on previous experience. Indeed, statistical prediction is central to many theories of nervous system function, including reinforcement learning (9, 10), optimal sensory-motor control (2, 11), decision making (5, 8, 12, 13), and efficient coding of sensory information (14–17). Optimal statistical predictions reflect not only the expected value (i.e., mean) of upcoming events but also the uncertainty (i.e., entropy) of those values (18, 19). The ability to form statistical predictions of sequential events suggests that neural circuits underlying perception and production of those sequences generate activity that reflects these predictions (4, 8–10, 12, 16, 17, 20–25). In contrast to sequential neuronal activation during production or replay of deterministic sequences (3, 24, 26–28), experimental demonstration of circuits exhibiting predictive activity shaped by the specific, long-term experienced statistical structure of upcoming events are rare. Here, we use the Bengalese finch (Bf) as a model system to test the hypothesis that the dynamics of internally generated neural activity are predictive of specific events and are shaped by the long-term, learned statistics of the timing and sequencing of those events.

Songbirds provide an excellent model for investigating neural processing involved in perception and production of complex sequential behavior. Birdsong is a learned behavior in which each bird produces a unique sequence of categorical acoustic elements,

termed syllables. Fig. 1*A* displays an example song segment from one Bf, represented here as a spectrogram (color intensity indicates energy at a given frequency over time). The sequencing of this bird's syllables, derived from multiple renditions of song, is summarized in Fig. 1*B*, which displays a transition diagram showing the (forward) conditional probability of transitioning from any given syllable to any other syllable. For the Bf, sequencing of syllables can be variable; multiple syllables may, in principle, transition to a given syllable (convergence points, e.g., “g” in Fig. 1*B*), and a particular syllable may, in principle, transition to multiple syllables (divergence points, e.g., “h” in Fig. 1*B*). Despite the variability in sequence production, Bf sequencing is not random, as the probabilistic structure of syllable transitions is reproducible across months (29). In addition to variability in sequences, silent intervals between syllables associated with a given transition, intersyllable gaps (ISGs), also exhibit reproducible stochastic structure that can differ across different transitions and may be linked to underlying sequence generation mechanisms (30, 31). Fig. 1*C* displays Gaussian fits to all gap distributions derived from songs for this bird. As is typical for Bfs, the gap distributions exhibited a range of mean values. Thus, song is a learned behavior with diverse statistics associated with both timing and sequencing (temporal-sequential statistics) of syllables.

The brain regions responsible for perception, acquisition, and production of song are well studied. Sensory-motor nucleus HVC (acronym used as proper name) transmits much of the descending motor control for the temporal structure of learned song (3, 32–35) and contributes to syllable sequencing (32, 34, 36). Interestingly, HVC neurons also respond more strongly to playback of the bird's

Significance

The neural basis of prediction is poorly understood. Here, utilizing the probabilistic song sequences of the Bengalese finch, we recorded neural activity following the termination of auditory playback of an individual's produced song sequences. We demonstrate that induced neural activity is predictive of the timing and identity of individual syllables. Consistent with the intuition that more uncertain sequences should give rise to weaker predictions, we find that the uncertainty of upcoming syllables modulates the strength of neural predictions. The utility of such statistical predictions for a variety of neural computations suggest that similar properties of neural dynamics may be a general feature of sensory and motor circuits.

Author contributions: K.E.B. and M.S.B. designed research; K.E.B. performed research; K.E.B. analyzed data; and K.E.B. and M.S.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

¹To whom correspondence should be addressed. Email: kebouchard@lbl.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1606725113/-DCSupplemental.

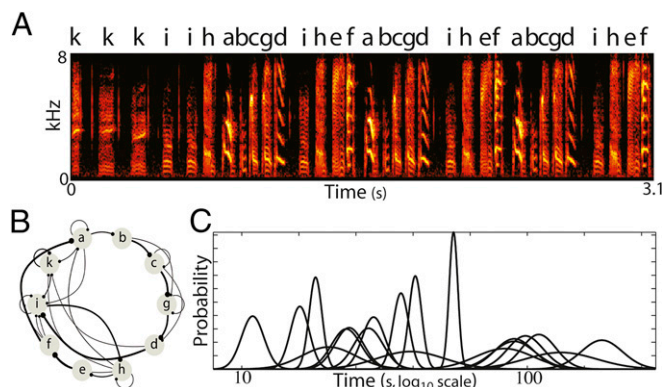


Fig. 1. Temporal-sequential structure of birdsong. (A) Spectrogram (power at frequency vs. time) of a Bf song segment, with letter labels assigned to unique syllables. Adult Bf song typically consists of 5–12 unique syllables produced in probabilistic sequences. (B) Forward transition diagram of song syntax, compiled from many renditions of the song in A. Each node corresponds to a unique syllable from the bird's repertoire. Edges between nodes denote transitions between syllables obtained from a large corpus of songs (dots denote transition destination). The width of each edge corresponds to the observed probability of transitioning from a syllable to any other syllable. (C) Intersyllable gap (ISG) probability distributions for all transitions in this bird. Each distribution is presented as the Gaussian fit to the \log_{10} ISG data.

own song (BOS) than to manipulated versions of BOS, such as songs with reverse syllable ordering (37–39). Indeed, previous work in Bfs has shown that HVC auditory responses exhibit long-time, probability-dependent integration that is shaped by the statistics of produced song sequences (40). We designed a series of experiments to examine whether neural activity in sensory-motor nucleus HVC persists after the termination of playback of syllable sequences, whether this activity is predictive of the timing and sequencing of produced syllable transitions, and whether the strength of predictions is modulated by the uncertainty of upcoming syllables.

Results

We played back auditory stimuli to 12 sedated adult Bfs while recording from HVC (*Materials and Methods*). We first characterized activity elicited by playback of BOS and reversed versions of BOS. Fig. 2 A and B shows example oscillograms of a BOS stimulus (BOS segment is excised from a longer bout of singing) and the temporally reversed BOS stimulus (rBOS), a control stimulus with the same spectral and amplitude content of BOS. The evoked auditory responses at one site to BOS and rBOS are displayed below the corresponding stimuli (mean \pm SEM, $n = 25$ trials). The mean evoked spike rate relative to the mean baseline activity (evoked/baseline) was significantly greater for BOS evoked auditory responses than rBOS auditory responses (across $n = 83$ sites from 12 birds: 5.15 ± 0.62 vs. 1.1 ± 0.53 , mean \pm SEM, $***P \leq 10^{-10}$, paired t test on \log_{10} data; across $n = 12$ birds: 4.45 ± 0.59 vs. 1.46 ± 0.16 , mean \pm SEM, $***P = 0$, Wilcoxon signed-rank test on \log_{10} data). Thus, as reported previously, Bf HVC auditory responses to BOS were much larger than auditory responses to rBOS (35, 41, 42).

If internally generated activity in HVC is predictive of future syllables, then this activity should persist after the termination of the BOS stimulus. Indeed, playback of BOS induced poststimulus activity that persisted for several hundred milliseconds after stimulus offset. Examples of poststimulus activity are displayed at the *Bottom* of Fig. 2 A and B for BOS and rBOS stimuli, where we magnify the poststimulus activity from directly above (corresponding time periods demarcated in gray). BOS-induced poststimulus activity decayed roughly exponentially after stimulus offset, whereas rBOS did not induce similarly temporally extended poststimulus activity. Across all sites, BOS-induced poststimulus activity decayed

to baseline levels after ~ 300 ms (exponential fit: $R^2 = 0.22$, $P < 10^{-4}$, $n = 83$ sites, *Materials and Methods*) and dipped slightly below baseline for ~ 700 ms before returning to approximate baseline levels (Fig. 2C). In contrast, rBOS poststimulus activity hovered around baseline and did not exhibit a systematic dependence on time (exponential fit: $R^2 = 0.008$, $P > 0.4$, $n = 83$ sites). Qualitatively, the poststimulus activity following BOS playback exhibited an initial (0–30 ms following BOS offset; Fig. 2D) increase in residual power (low-pass-filtered residuals from best-fit exponential at each site; *Materials and Methods*) followed by more subtle fluctuations at the average times at which syllables would have occurred in produced song (gray shaded regions, Fig. 2D). Poststimulus activity that could exhibit nonmonotonic decay following BOS playback was also apparent in both multiunit and single-unit activity recorded at single sites (e.g., Fig. 4A, red trace, and Fig. S1). The presence of such poststimulus activity following BOS playback raises the possibility that ongoing HVC activity carries a prediction about upcoming events in song. Here, we tested whether such predictions are present for the timing and identity of syllables that would normally follow the segments of BOS that were played back to birds.

Temporal Tuning for Syllable Transitions Reflects the Expected Timing of Produced Intersyllable Gaps. If the dynamics of HVC neural activity are structured by the temporal-sequential statistics of produced song, then auditory responses to sequences of syllables

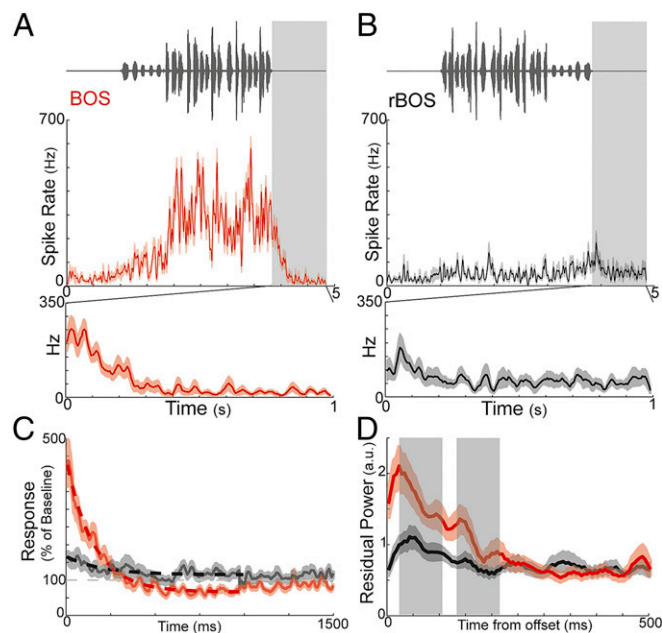


Fig. 2. Poststimulus activity in HVC induced by BOS playback. Oscillograms (*Top*) of BOS (A) and rBOS (B), with evoked HVC auditory responses at one site (*Middle*: mean \pm SEM, $n = 25$ trials) and induced poststimulus activity demarcated with gray boxes expanded at *Bottom*. HVC auditory responses evoked by BOS were larger than those evoked by rBOS, and the corresponding induced poststimulus activity was also larger. (C) Average poststimulus activity across HVC sites, expressed as percent above baseline, induced by auditory playback of BOS (red) and rBOS (black) stimuli (mean \pm SEM; $n = 83$ sites from 12 birds). Dashed red and black lines are best-fitting exponential decay. Gray dashed line corresponds to baseline activity. (D) Average power (5–25 Hz) of residuals from exponential fits of poststimulus activity across HVC sites induced by auditory playback of BOS (red) and rBOS (black) stimuli (mean \pm SEM; $n = 83$ sites from 12 birds). Gray boxes demarcate average duration of syllables, separated by median gap. BOS-induced residuals were significantly greater than rBOS-induced residuals over the first 250 ms of poststimulus activity (across $n = 83$ sites from 12 birds; BOS, 1.50 ± 0.14 , vs. rBOS, 0.9 ± 0.09 ; mean \pm SEM; $P \leq 10^{-4}$, paired t test; across $n = 12$ birds; BOS, 1.40 ± 0.19 , vs. rBOS, 0.86 ± 0.12 ; mean \pm SEM; $P \leq 10^{-3}$, Wilcoxon signed-rank test).

should be tuned to the specific timing of those sequences during song production. However, the degree to which the timing between syllables presented during auditory playback modulates HVC auditory responses is poorly understood. A previous study reported responses of a single HVC neuron that decayed monotonically with increasing time between syllables (39). Such a monotonically decreasing response might be explained simply if the response to delayed syllables in the sequence were superimposed on gradually decaying activity elicited by the preceding sequence. In contrast, we hypothesized that persistent HVC activity contains a more specific prediction of the timing at which subsequent syllables of song normally occur. To test for such a prediction, we played back segments of the BOS (three to five syllables) and varied the time between the last two syllables across a range that contained durations that were both shorter and longer than the mean ISG for that transition. For example, if a bird produced sequence “abcde” (Fig. 3A, *Top*), we played back the sequence “abcde,” and measured responses to “e” as a function of the gap between “d” and “e.” We asked whether responses to the delayed syllable depended non-monotonically on the delay at which the syllable was played, and whether there was any relationship between the delay at which the strongest responses to the terminal syllable occurred and the duration of the ISG that normally preceded that syllable in song.

Fig. 3B illustrates temporal tuning for one transition. The top panel shows the response at one site to the last syllable as a function of the temporal offset (dt) from the previous syllable. The bottom panel shows the average response across four sites. The probability distribution of produced ISGs for this transition is presented in the bottom panel. For this transition, we found that the peak in temporal tuning occurred at an interval of 30 ms, which was very close to the mean of the produced ISG distribution (33 ms). Fig. 3C presents results from a different bird for a transition where the mean of the produced ISGs (*Bottom*) was much longer. The peak in temporal tuning was again well aligned to the peak in the produced ISG distribution (*Bottom*).

Across all 27 sites in three birds at which a total of six transitions were tested, peak temporal tuning was well aligned to the mean duration of produced gaps. Fig. 3D summarizes temporal tuning for all sites as function of time between the last two syllables relative to the mean produced ISG for associated transitions (Δt). The average temporal tuning function ($n = 27$ sites) exhibited a peak near $\Delta t = 0$, indicating that peak responses occur near the mean of produced ISG distributions. Indeed, across all sites, there was a strong positive correlation between the peak of the temporal tuning curve and the mean of the produced ISG distribution (Fig. 3E, gray, $n = 27$ sites at which six transitions were played back; $R^2 = 0.67$, $P < 10^{-10}$; red, responses averaged across sites for $n = 6$ transitions from three birds; $R^2 = 0.93$, $P < 0.002$). However, peak tuning was delayed by a small but significant amount (12 ± 3 ms) relative to the produced mean ISG for a transition ($P < 10^{-3}$, t test, $N = 27$). As there is no physical energy in the stimulus during gaps to drive neural activity, the observed temporal tuning must result from dynamics of internally generated neural activity induced by preceding sequences. These results are consistent with the hypothesis that HVC dynamics are predictive for the expected timing of upcoming syllables.

Poststimulus Neural Activity Is Predictive of Upcoming Syllable Identity for Low-Uncertainty Transitions. The observation of temporal tuning shaped by the timing of upcoming syllables suggests that HVC activity might also predict the identity of upcoming syllables. We tested the specificity of sequence-induced poststimulus activity for the next syllable in a sequence. Bf song contains transitions where the upcoming syllable is highly certain, and transitions where the upcoming syllable is uncertain. We quantified syllable transition uncertainty as conditional entropy: lower entropy corresponds to lower uncertainty in the identity of upcoming syllables. Transitions exhibited a bimodal distribution of entropies, which was perfectly separable (i.e., completely nonoverlapping, area under receiver

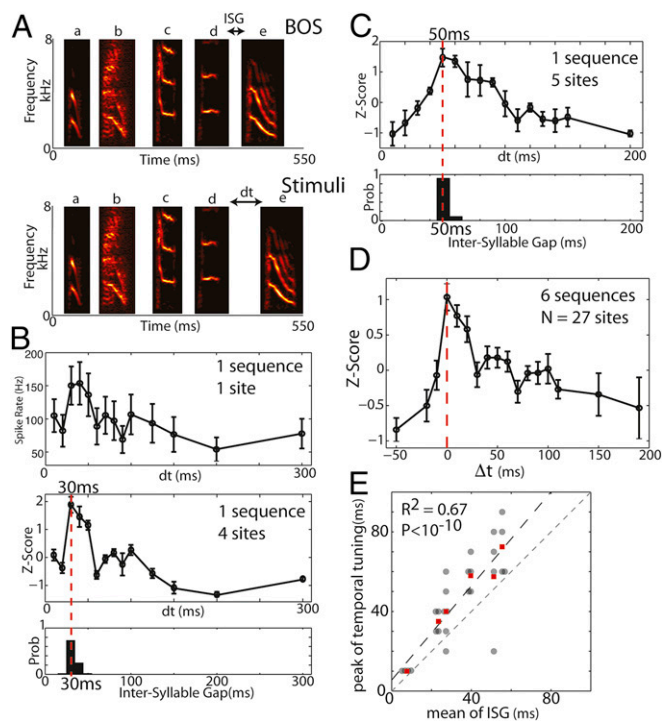


Fig. 3. Temporal tuning of HVC auditory responses for syllable transitions reflects the expected timing of produced intersyllable gaps (ISGs). (A) Illustration of stimulus used to determine temporal tuning. To investigate whether the temporal tuning of HVC auditory responses to a syllable is matched to the statistics of song production, we presented sequences of syllables and varied the time between the last two syllables (*Bottom*) across values that were both shorter and longer than the average of naturally produced ISGs (*Top*). (B, *Top*) Average spike rate in response to the last syllable as a function of the temporal offset from the previous syllable (dt) at one site (mean \pm SEM; $n = 20$). (*Middle*) Average responses across sites (z scores; mean \pm SEM; $n = 4$ sites). (*Bottom*) The probability distribution of ISGs for this transition produced by the bird during singing. The temporal tuning exhibited a peak (time of peak demarcated by dashed red line) at 30 ms that was well aligned to the mean of the produced ISG's for that transition (32 ms). (C) Another example showing the response as a function of dt (*Top*: z scores; mean \pm SEM; $n = 5$ sites) for a sequence of syllables with a longer produced ISG (*Bottom*). Here, the mean produced gap duration was 53 ms, and the temporal tuning exhibited a peak at 50 ms (dashed red line). (D) Mean z scores as a function of the difference between the presented ISG and the mean produced ISG for six transitions at 27 sites from three birds [$\Delta t = dt - \text{mean}(\text{ISG})$]. Black curve: mean \pm SEM, $n = 27$ sites. Vertical dashed red line demarcates $\Delta t = 0$. (E) Location of peaks in temporal tuning curves for individual sites and transitions as a function of the mean of the produced ISG distributions. There was a strong correlation between the locations of the peaks of temporal tuning curves and the means of produced ISGs for the corresponding transitions (black dots: individual sites, small horizontal shift added to overlapping points; dashed black line is best linear fit, $R^2 = 0.67$, $P < 10^{-10}$, six transitions played at $n = 27$ sites; red squares: mean across sites for a transition, $n = 6$ transitions from three birds; $R^2 = 0.93$, $P < 0.002$; dashed gray line is unity).

operating characteristic curve = 1) at 0.75 bits; we therefore split the data into sequences ending in “low entropy” (< 0.75 bits) or “high entropy” (> 0.75 bits) transitions. We hypothesized that poststimulus activity induced by sequences ending in low-entropy transitions would be predictive of upcoming syllables. To test this hypothesis, we played back short sequences of syllables (three to six syllables) from BOS, as well as individual syllables in isolation (in total, 88 sequences were played back, 44 low entropy and 44 high entropy) at 18 multiunit sites in seven birds, and responses to identical stimuli were averaged across multiple sites within a bird. For the example in Fig. 4A, syllable “e” always followed the sequence “abcd” during song production, and so the transition from “d–e” was low entropy. Fig. 4A illustrates the average auditory

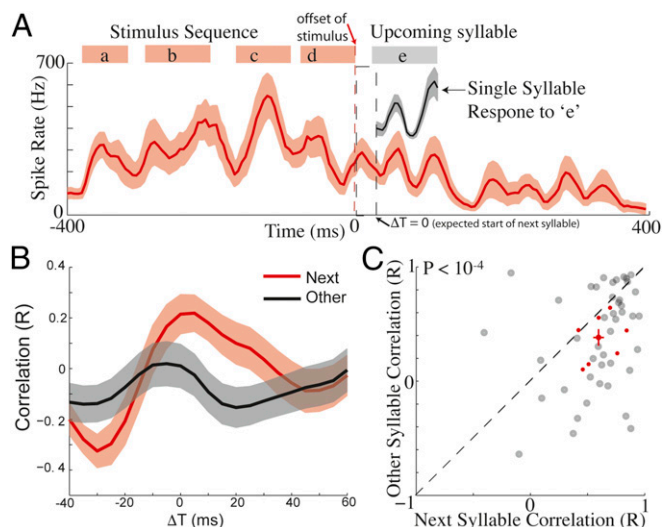


Fig. 4. Poststimulus activity is predictive of upcoming syllables for low uncertainty transitions. (A) Poststimulus activity induced by playback of a BOS sequence was similar to the auditory response to the next syllable in the sequence. The spike rate evoked by playback of the sequence “abcd” (syllable timing in stimulus demarcated by red boxes), as well as the induced poststimulus activity (stimulus offset is indicated with red dashed line). The single-syllable response to “e” is presented above the period when it would have occurred (indicated by gray box). $\Delta T = 0$ is the average expected time of next syllables. Data are presented as mean \pm SEM across trials. (B) Time course of correlation coefficients comparing poststimulus activity to evoked single-syllable responses for the next syllable in the sequence (red) and other syllables in the birds repertoire (black) (mean \pm SEM; $n = 44$ single-syllable comparisons from 18 sites in seven birds). (C) Scatter plot contrasting maximum correlation coefficients comparing poststimulus activity to evoked single-syllable responses for next syllables and other syllables. Black circles are individual data points, and red cross is mean \pm SEM for the 44 unique syllable transitions; red points are within-bird averages for seven birds with one to four recording sites, and four to nine unique syllable transitions per bird.

response evoked by playback of sequence “abcd,” and the induced poststimulus activity. The average single-syllable response to “e” is presented above the period when it would have occurred had the sequence “abcde” been presented. Poststimulus activity induced by “abcd” was qualitatively similar to the response evoked by syllable “e,” raising the possibility that HVC activity following termination of the sequence “abcd” anticipates, or “predicts,” the specific temporal structure of activity elicited by the next syllable in song (in this case, “e”).

We quantified similarity between poststimulus activity and single-syllable response waveforms evoked by the next syllable(s) in the sequence using a sliding covariance analysis (*Materials and Methods*). This measures the similarity between single-syllable responses and poststimulus activity of the same duration. To determine the time course of these correlations, we carried out this analysis over a range of temporal lags spanning the expected onset of the next syllable(s) (*Materials and Methods*). To determine selectivity of poststimulus activity for the identity of the next syllable(s), we compared next-syllable correlations with correlations between poststimulus activity and responses to other syllables in the bird’s repertoire (i.e., those that did not follow the presented sequences during song production). For a given sequence, if there were multiple “next” syllables, each contributed a data point, whereas the correlations for all “other” syllables were averaged together.

Across all low-entropy transitions, poststimulus activity was more similar to responses for next syllable(s) than for other syllables. Fig. 4B presents the average correlation (relative to the mean expected timing, $\Delta T = 0$) comparing poststimulus activity to evoked single-syllable responses for the next syllable(s) in the

sequence and other syllables. Correlations for next syllables (red) were greater than the correlations for other syllables (black). Relative to the expected timing ($\Delta T = 0$), the peak of the average correlation for next syllables occurred with a delay of 7 ms, and the median of peaks for individual correlations was delayed by 9 ms (Fig. 4B, $N = 44$). These delays were similar to the delay observed for temporal tuning (12 ms; Fig. 3). The maximum correlation coefficients for next syllables (extracted between -20 and 40 ms after the offset of the stimulus sequence) were significantly greater than for other syllables (Fig. 4C, $P < 10^{-4}$, Wilcoxon signed-rank test; $n = 44$ syllables). These results indicate that the dynamics of post-stimulus neural activity in HVC are predictive of upcoming syllable identity.

Uncertainty of Produced Sequences Modulates the Degree to Which Poststimulus Activity Is Predictive of Upcoming Syllables. We next used the diversity of transition uncertainties found in Bf song sequences to examine whether predictive strength of HVC post-stimulus activity is modulated by the uncertainty of upcoming syllables. We hypothesized that poststimulus activity induced by sequences ending in uncertain transitions would be less predictive of upcoming syllables than for sequences ending in more certain transitions. Fig. 5A presents for high-entropy transitions (>0.75 bits) average correlations between poststimulus activity and evoked single-syllable responses for the next syllable(s) in the sequence (red) and the other syllables (black). Here, maximum correlation coefficients were statistically indistinguishable between next and other syllables (Fig. 5B, $P > 0.3$, Wilcoxon signed-rank test, $n = 44$ syllables; red cross is mean \pm SEM). Thus, for high-entropy (i.e., uncertain) transitions, unlike the case for low-entropy transitions, we did not detect selectivity of post-stimulus activity for upcoming syllable identity.

To directly examine how predictive strength differed between low- and high-entropy sequences, we quantified predictive strength as the difference between maximum next-syllable correlation coefficients and maximum other-syllable correlation coefficients (ΔR). The histogram at Bottom of Fig. 5C plots transition entropies for all tested sequences. The entropies formed a bimodal distribution, with a large margin between the low-entropy (red) and high-entropy (black) modes. A paired comparison of predictive strength for low- and high-entropy transitions indicated that predictive strength was larger for low-entropy transitions (0.224 ± 0.071 ; mean \pm SEM;

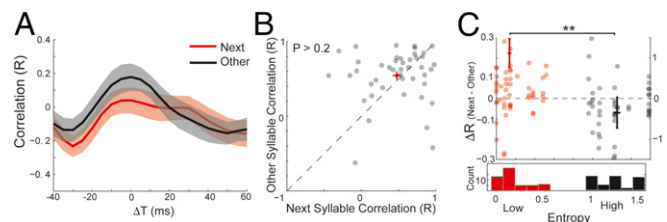


Fig. 5. Uncertainty of produced sequences modulates the degree to which poststimulus activity is predictive of upcoming syllables. (A) Time course of correlation coefficients comparing poststimulus activity to evoked single syllable responses for the next syllable in the sequence (red) and other syllables in the birds repertoire (black) (mean \pm SEM; $n = 44$). $\Delta T = 0$ is the average expected time of next syllables. (B) In contrast to results for low-entropy transitions, for high-entropy transitions we found that maximum correlation coefficients were statistically indistinguishable between next and other syllables ($P > 0.3$, Wilcoxon signed-rank test, $n = 44$ syllables; black circles are individual syllables, and red is mean \pm SEM). (C) Transition uncertainty modulates the degree to which poststimulus activity is predictive of upcoming syllables. Difference in correlation coefficients (ΔR) between next and other syllables was significantly larger for low-entropy transitions than for high-entropy transitions (** $P = 0.006$, Wilcoxon signed-rank test; $n = 44$ syllables for both).

thick red line) than for high-entropy transitions (-0.070 ± 0.076 ; mean \pm SEM; thick black line) (Fig. 5C, numerical values on left ordinate, $**P = 0.006$, Wilcoxon signed-rank test, $n = 44$ sequences for both low and high entropy). Furthermore, average predictive strength was greater for low-entropy transitions than for high-entropy transitions in five out of seven individual birds. These results were further confirmed by regressing ΔR against transition entropy for individual transitions ($R = -0.28$, $P = 0.006$, $n = 88$; Fig. 5C, colored dots, numerical values indicated on right ordinate). There were no significant differences in responses to the last syllable between the high- and low-entropy sequences (Fig. S2). These results demonstrate that transition entropy modulates the predictive strength of poststimulus activity in HVC; low-entropy transitions induced poststimulus activity that was more similar to the single-syllable responses for next syllables than other syllables, whereas high-entropy transitions did not. Correspondingly, predictive strength (ΔR) was significantly greater for low-entropy than for high-entropy sequences.

Discussion

An individual's ability to form statistical predictions of sequential events implies that neural circuits must be able to leverage information about past experience to predict future events (1–4, 8–10, 12, 16, 17, 20–25). We used playback of birdsong to investigate whether and how activity in sensorimotor nucleus HVC encodes information about song sequence statistics to predict upcoming sensorimotor events. We found that song playback induces ongoing activity in HVC predictive of the timing and identity of upcoming syllables, and that the strength of this predictive activity depends on the certainty with which upcoming syllables occur.

A previous study examining the spectrotemporal selectivity of HVC auditory responses reported one neuron for which responses to terminal syllables decreased monotonically as the delay between syllables was progressively increased (39). However, that study used stimuli with timing between syllables that only exceeded the produced ISGs and therefore was not designed to sample the range of timing delays required to detect a correspondence between temporal tuning and produced gap durations. In contrast, we found that responses to playback of syllable sequences did not decay monotonically as a function of delay, but rather were greatest when those sequences were presented with ISGs that matched those normally produced during singing. Thus, in HVC, the peak of temporal tuning is correlated with the expected value of the produced ISG distributions for specific transitions. Because there is no physical energy in the stimulus during the time between syllables, this temporal tuning for specific syllable transitions must arise from dynamics of internally generated neural activity, and is consistent with the hypothesis that the neural dynamics in HVC are predictive of the temporal statistics of song.

We also tested whether poststimulus activity induced by playback of syllable sequences resembled single-syllable responses evoked by the next syllables in the sequence. Previous studies in the zebra finch characterized “temporal combination-selective” cells, in which the response of a neuron to playback of two normally ordered syllables is supralinear relative to responses to the syllables in isolation, or in abnormal order (37, 39). Similarly, in the Bf, HVC responses exhibit long-time, probability-dependent sequence integration (42). These forms of temporal integration suggest that neural activity following offset of a particular sequence can persist until the presentation of the next syllable in the sequence. Moreover, enhanced responses to normally ordered syllables suggest that such neural activity includes some prediction of the identity of upcoming syllables. Indeed, playback of normally sequenced elements of the BOS can evoke patterns of activity that resemble those present during singing (43). Here, we found that, for low-uncertainty transitions, poststimulus activity following a

sequence of syllables resembles the temporal structure of activity that would normally be elicited by playback of the next syllable in the sequence. Hence, when the next syllable is predictable with greater certainty, neural responses are predictive of that syllable. Moreover, in agreement with the intuition that more uncertain transitions should give rise to weaker predictions about upcoming syllables, poststimulus activity induced by low-entropy transitions was more selective for upcoming syllables than poststimulus activity induced by high-entropy transitions.

Together, these results demonstrate that the dynamics of internally generated neural activity are predictive of the identity of upcoming syllables, and that the uncertainty of a given transition modulates the strength of this prediction. This predictive activity may, in part, contribute to long-time probability-dependent sequence integration (42), by providing a “trace” of the preceding sequence and a prediction of the upcoming syllable. Specifically, incoming sensory signals from a syllable may be integrated with “reverberating” activity evoked by the preceding sequence. Interestingly, distorted auditory feedback during singing has a larger effect on sequence production and HVC neural activity at more entropic transitions (35). This could reflect differences in the strength of “prediction” for upcoming syllables in ongoing activity, which might promote the corresponding, highest probability transitions. Given the strong sensory-motor correspondences in HVC, the entropy-dependent, predictive poststimulus activity demonstrated here could be a basis for biasing upcoming transitions and for the differential effects of distorted auditory feedback on higher vs. lower entropy transitions.

Although our results derive from recordings in HVC, the mechanisms that underlie generation of predictive activity are likely distributed across multiple nuclei. The network mechanisms underlying syllable sequencing include contributions from local microcircuits in HVC (31–36, 44), refferent activity through extended brain circuits (e.g., HVC \rightarrow RA \rightarrow ... \rightarrow HVC) (45), and activation due to sensory feedback (29, 35, 46). Our observation that the dynamics of poststimulus activity were slowed by ~ 10 ms relative to the expected timing of song [for both temporal tuning (Fig. 3 D and E) and peak predictive strength (Fig. 4B)] also suggests that the song generation network is in a different state during singing, perhaps resulting from different neuromodulatory tone (47), or from additional excitatory inputs that are not activated during playback experiments (32–35, 45).

Our findings have similarities to observations in other systems in which neural circuitry encodes information about the statistics of experience. For example, in the visual system of mice, spatiotemporal patterns of spontaneous neural activity resemble those evoked by visual stimuli for a period immediately following stimulus presentation (48). Similarly, spontaneous activity in the experienced ferret visual system can exhibit greater similarities to activity evoked by natural movies than in naïve ferrets, suggesting that spontaneous activity reflects prior natural sensory experience (49). Moreover, the hippocampus in many species has been implicated in both the encoding and recall of memory episodes, and spontaneous hippocampal neural dynamics have been associated with sequence prediction (23, 24, 26, 28). These observations and others indicate that prior experiences can shape predictive neural activity, as we have observed here. Our results additionally indicate that such predictive neural activity reflects the strength of statistical regularities of specific events in an individual's long-term experiences. We have provided converging lines of evidence indicating that playback of sequences of syllables induces internally generated neural activity, the dynamics of which are predictive of the learned temporal-sequential statistics of specific sensory-motor events from long-term exposure. Our results derive from a sensorimotor circuit specialized for perception and production of song. Hence, it remains to be determined

whether there is similar encoding of stimulus specific predictions in other systems. However, the utility of statistical predictions for a variety of neural computations and the ability of Hebbian learning to engrain these statistics in neural circuits (50) suggest that similar properties of neural dynamics may be a general feature of sensory and motor circuits.

Materials and Methods

Twelve male Bfs (age, >110 d) were used in this study. During experiments, birds were housed individually in sound-attenuating chambers (Acoustic

Systems), and food and water were provided ad libitum. Acute electrophysiology procedures, as previously described (42), were performed in accordance with established animal care protocols approved by the University of California, San Francisco, Institutional Animal Care and Use Committee. Birds used in this study also contributed data to other studies (42, 46, 50). See *SI Materials and Methods* for further details.

ACKNOWLEDGMENTS. This work was supported by the Howard Hughes Medical Institute, National Institutes of Health Grant DC006636 (to M.S.B.), National Science Foundation Grant IOS0951348 (to M.S.B.), and a National Science Foundation predoctoral award (to K.E.B.).

- Lashley KS (1951) The problem of serial order in behavior. *Cerebral Mechanisms in Behavior*, ed Jeffress LA (Wiley, New York), pp 112–147.
- Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination. *Nat Neurosci* 5(11):1226–1235.
- Hahnloser RHR, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419(6902):65–70.
- Bouchard KE, Chang EF (2014) Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. *J Neurosci* 34(38):12662–12677.
- Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006) The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113(4):700–765.
- Peña M, Bonatti LL, Nespor M, Mehler J (2002) Signal-driven computations in speech processing. *Science* 298(5593):604–607.
- Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science* 274(5294):1926–1928.
- Yang T, Shadlen MN (2007) Probabilistic reasoning by neurons. *Nature* 447(7148):1075–1080.
- Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat Neurosci* 1(4):304–309.
- Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299(5614):1898–1902.
- Wiener N (1961) *Cybernetics* (The MIT Press, Cambridge, MA).
- Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. *Science* 304(5678):1782–1787.
- Gold JJ, Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci* 30:535–574.
- Barlow HB (1961) Possible principles underlying the transformation of sensory messages. *Sensory Communication*, ed Rosenblith WA (MIT Press, Cambridge, MA), pp 217–234.
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381(6583):607–609.
- Hosoya T, Baccus SA, Meister M (2005) Dynamic predictive coding by the retina. *Nature* 436(7047):71–77.
- Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2(1):79–87.
- Knill DC, Pouget A (2004) The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci* 27(12):712–719.
- Bach DR, Dolan RJ (2012) Knowing how much you don't know: A neural organization of uncertainty estimates. *Nat Rev Neurosci* 13(8):572–586.
- DeLong KA, Urbach TP, Kutas M (2005) Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat Neurosci* 8(8):1117–1121.
- Diba K, Buzsáki G (2007) Forward and reverse hippocampal place-cell sequences during ripples. *Nat Neurosci* 10(10):1241–1242.
- Leaver AM, Van Lare J, Zielinski B, Halpern AR, Rauschecker JP (2009) Brain activation during anticipation of sound sequences. *J Neurosci* 29(8):2477–2485.
- Lisman J, Redish AD (2009) Prediction, sequences and the hippocampus. *Philos Trans R Soc Lond B Biol Sci* 364(1521):1193–1201.
- Pastalkova E, Itskov V, Amarasingham A, Buzsáki G (2008) Internally generated cell assembly sequences in the rat hippocampus. *Science* 321(5894):1322–1327.
- Zhou X, de Villers-Sidani E, Panizzutti R, Merzenich MM (2010) Successive-signal biasing for a learned sound sequence. *Proc Natl Acad Sci USA* 107(33):14839–14844.
- Davidson TJ, Kloosterman F, Wilson MA (2009) Hippocampal replay of extended experience. *Neuron* 63(4):497–507.
- Ikegaya Y, et al. (2004) Synfire chains and cortical songs: Temporal modules of cortical activity. *Science* 304(5670):559–564.
- Karlsson MP, Frank LM (2009) Awake replay of remote experiences in the hippocampus. *Nat Neurosci* 12(7):913–918.
- Warren TL, Charlesworth JD, Tumer EC, Brainard MS (2012) Variable sequencing is actively maintained in a well learned motor skill. *J Neurosci* 32(44):15414–15425.
- Glaze CM, Troyer TW (2006) Temporal structure in zebra finch song: Implications for motor coding. *J Neurosci* 26(3):991–1005.
- Jin DZ (2009) Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. *Phys Rev E Stat Nonlin Soft Matter Phys* 80(5 Pt 1):051902.
- Fee MS, Kozhevnikov AA, Hahnloser RHR (2004) Neural mechanisms of vocal sequence generation in the songbird. *Ann N Y Acad Sci* 1016:153–170.
- Long MA, Jin DZ, Fee MS (2010) Support for a synaptic chain model of neuronal sequence generation. *Nature* 468(7322):394–399.
- McCasland JS (1987) Neuronal control of bird song production. *J Neurosci* 7(1):23–39.
- Sakata JT, Brainard MS (2008) Online contributions of auditory feedback to neural activity in avian song control circuitry. *J Neurosci* 28(44):11378–11390.
- Long MA, Fee MS (2008) Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456(7219):189–194.
- Lewicki MS, Arthur BJ (1996) Hierarchical organization of auditory temporal context sensitivity. *J Neurosci* 16(21):6987–6998.
- Lewicki MS, Konishi M (1995) Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc Natl Acad Sci USA* 92(12):5582–5586.
- Margoliash D, Fortune ES (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J Neurosci* 12(11):4309–4326.
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495(7441):327–332.
- Nakamura KZ, Okanoya K (2004) Neural correlates of song complexity in Bengalese finch high vocal center. *Neuroreport* 15(8):1359–1363.
- Bouchard KE, Brainard MS (2013) Neural encoding and integration of learned probabilistic sequences in avian sensory-motor circuitry. *J Neurosci* 33(45):17710–17723.
- Dave AS, Margoliash D (2000) Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* 290(5492):812–816.
- Vu ET, Mazurek ME, Kuo YC (1994) Identification of a forebrain motor programming network for the learned song of zebra finches. *J Neurosci* 14(11 Pt 2):6924–6934.
- Ashmore RC, Wild JM, Schmidt MF (2005) Brainstem and forebrain contributions to the generation of learned motor behaviors for song. *J Neurosci* 25(37):8543–8554.
- Wittenbach J, Bouchard KE, Brainard MS, Jin DZ (2015) An adapting auditory-motor feedback loop can contribute to generating vocal repetition. arXiv:1501.527.
- Dave AS, Yu AC, Margoliash D (1998) Behavioral state modulation of auditory activity in a vocal motor system. *Science* 282(5397):2250–2254.
- Han F, Caporale N, Dan Y (2008) Reverberation of recent visual experience in spontaneous cortical waves. *Neuron* 60(2):321–327.
- Berkes P, Orbán G, Lengyel M, Fiser J (2011) Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331(6013):83–87.
- Bouchard KE, Ganguli S, Brainard MS (2015) Role of the site of synaptic competition and the balance of learning forces for Hebbian encoding of probabilistic Markov sequences. *Front Comput Neurosci* 9:92.