Published in final edited form as:

Nat Neurosci. 2011 March ; 14(3): 373–380. doi:10.1038/nn.2748.

Learning the microstructure of successful behavior

Jonathan D. Charlesworth^{1,2}, Evren C. Tumer¹, Timothy L. Warren^{1,2}, and Michael S. Brainard^{1,2}

¹W. M. Keck Center for Integrative Neuroscience and Department of Physiology, University of California, San Francisco, California 94143, USA

²Neuroscience Graduate Program, University of California, San Francisco, California 94143, USA

Abstract

Reinforcement signals indicating success or failure are known to alter the probability of selecting between distinct actions. However, successful performance of many motor skills, such as speech articulation, also requires learning behavioral trajectories that vary continuously over time. Here, we investigated how temporally discrete reinforcement signals shape a continuous behavioral trajectory, the fundamental frequency of adult Bengalese finch song. We provided reinforcement contingent on fundamental frequency performance only at one point in song. Learned changes to fundamental frequency were maximal at this point, but also extended both earlier and later in the fundamental frequency trajectory. A simple principle predicted the detailed structure of learning; birds learn to produce the average of the behavioral trajectories associated with successful outcomes. This learning rule accurately predicts the structure of learning at a millisecond time scale, demonstrating that the nervous system records fine-grained details of successful behavior and uses this information to guide learning.

Much prior research has focused on how reinforcement alters the probability of discrete actions, such as pressing one lever instead of another^{1–3}. In contrast, it remains unclear how reinforcement shapes the continuous trajectories of natural behaviors. This question is nevertheless central to understanding the role of reinforcement in learning, since behavioral success requires not only selecting appropriate actions but also controlling the detailed musculoskeletal trajectories that allow us to accomplish those actions. Consider for example a baseball pitcher attempting to strike out opposing batters. At a level of action selection, the pitcher must choose whether to throw a fastball or a curveball. In this context, reinforcement is known to increase the probability of selecting actions that result in more successful outcomes; the pitcher will throw more fastballs than curveballs if doing so results in more strikes. At a more detailed level, the placement and movement of the ball will depend on the continuous trajectory of the pitcher's hands, arms and legs. Qualitative analysis has suggested that reinforcement can influence such behavioral trajectories^{2,4}; the pitcher may modify the complex dynamics of his motion to favor trajectories that result in more strikes.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Author contributions All authors contributed to designing the experiments. E.C.T. developed the song acquisition and reinforcement delivery software. T.L.W., E.C.T. and J.D.C. performed the experiments and collected the data. J.D.C. developed the modeling and analysis techniques and prepared the manuscript, with assistance from T.L.W. and M.S.B.

However, the specific principles by which reinforcement shapes such continuous behavioral trajectories remain unclear.

In this study we characterized how temporally discrete reinforcement shapes a continuous trajectory of behavior on a millisecond time scale, near the temporal limits of neural control^{5–7}. We investigated the extent to which reinforcement that is contingent on behavior at a specific, discrete time, and thus causally related to behavior only at that time, influences the surrounding behavioral trajectory. We hypothesized that the nervous system records the structure of behavioral trajectories correlated with reinforcement and learns to adopt this structure into subsequent behavioral trajectories. This hypothesis predicts that learning will incorporate aspects of behavioral trajectories that were not causally linked to reinforcement but occurred by chance on reinforced trials. To test this hypothesis, we quantified reinforcement-driven learning in the trajectory of fundamental frequency for adult Bengalese finch song, a learned vocalization containing a sequence of multiple syllables each ~30–100ms in duration. Song is highly stereotyped in the absence of external perturbations and can be monitored with exceptional temporal resolution, allowing precise control of experimentally imposed reinforcement and fine-grained analysis of learning^{8–16}.

Bengalese finches can learn to modify the mean fundamental frequency of individual syllables in response to aversive reinforcement that is contingent on syllable-by-syllable variation in fundamental frequency⁸. The mean fundamental frequency of a specific syllable is highly stereotyped in the absence of external perturbations, yet exhibits subtle variation (coefficient of variation of 1–2%) that can be used for learning in the presence of reinforcement signals. Previous reports demonstrate that applying aversive reinforcement signals (~50ms bursts of white noise) contingent on variation in the fundamental frequency of a targeted syllable elicits rapid adaptive changes in the mean fundamental frequency of that syllable^{8,17}. These reports monitored a single value of fundamental frequency for each syllable rendition to determine whether aversive reinforcement would be applied to that syllable and to quantify learned changes to fundamental frequency^{8,17}.

To study learning in a more detailed fashion, we provided birds with discrete reinforcement that was contingent on performance at precisely timed points within individual syllables, and measured the microstructure of learning, the learned change in the continuous trajectory of fundamental frequency at a millisecond time scale. We found that a simple rule accurately predicted the microstructure of learning: birds learned to produce the average of the fundamental frequency trajectories that successfully escaped aversive reinforcement. These results indicate that the structure of temporal correlations in behavior can shape learning at a millisecond time scale and provide a simple account for the structure of learned behavior that develops in response to discrete reinforcement.

Results

We hypothesized that temporal correlations in the structure of successful behavioral variation determine the structure of learning. If so, it should be possible to predict the detailed structure of learning from the pattern of behavioral variation that avoids aversive

reinforcement. To test this principle, we conducted four sets of experiments that delivered aversive reinforcement to distinct subsets of behavioral variation.

Precise reinforcement yields temporally specific learning

First, we elicited learning using reinforcement signals contingent on fundamental frequency at a single, precise time in specifically targeted syllables of adult Bengalese finch song. We quantified learning by measuring the temporal trajectory of fundamental frequency for each rendition of the targeted syllable. To deliver reinforcement with millisecond precision, we developed software that recognized a specific time in a targeted syllable of adult Bengalese finch song (the “contingency time”, Fig. 1a, blue arrows) and triggered a burst of white noise (aversive reinforcement) contingent on fundamental frequency performance at that time (e.g. Fig. 1a, point A2). Relative to syllable onset, the median standard deviation of the contingency time was 7.9ms in syllables with 50ms mean duration (**see Methods**). Consistent with previous work⁸, selective delivery of aversive reinforcement to low fundamental frequency variants elicited adaptive increases in mean fundamental frequency at the contingency time (Fig. 1b), whereas delivery to high fundamental frequency variants induced decreases in fundamental frequency. In a given experiment, feedback was contingent on fundamental frequency at a specific time in one syllable of one bird's song and was designed to drive either upward or downward learning (n=28 experiments for 23 syllables in 21 birds). To simplify presentation, we have plotted data so that the adaptive direction is always upwards.

Qualitatively, reinforcement contingent on fundamental frequency at a single time drove conspicuous changes in the shape of the fundamental frequency trajectory for targeted syllables. Syllables are categorized as ‘flat harmonic stacks’ if the fundamental frequency remains relatively constant and as ‘frequency modulated sweeps’ if the fundamental frequency systematically falls (or rises) over the duration of the syllable. For flat harmonic stacks, reinforcement of higher (or lower) fundamental frequency variants did not simply shift the fundamental frequency of the entire syllable upwards or downwards. Rather, changes in fundamental frequency were most pronounced around the contingency time. In extreme cases, this could convert a flat harmonic stack into a syllable better described as a frequency modulated sweep (Supplementary Fig. 1). Such qualitative changes in response to temporally precise reinforcement suggest that learning does not simply reflect the adjustment of a single parameter (i.e. the mean fundamental frequency) for the entire syllable. Correspondingly, this observation raises the question of what principle guides the structure of learned changes to the fundamental frequency trajectory.

To investigate the detailed structure of learning, we quantified fundamental frequency trajectories with millisecond resolution (Fig.1c) using a short-time Fourier transform¹⁸ (**see Methods**). The learned trajectory (the microstructure of learning) was defined as the change in the fundamental frequency trajectory over a block of reinforcement trials. To assess relative differences in the trajectory of learning around the time of reinforcement, learned trajectories for each experiment were normalized to maximal learning in each experiment.

This quantification of learning confirmed that maximal learning was localized at the contingency time. We investigated this phenomenon by performing three experiments in a

single syllable, each with a different contingency time. After each experiment, we allowed fundamental frequency to recover back to its original value before the next experiment. Varying the contingency time in this fashion elicited distinct patterns of learning, but in each experiment, learning was maximal near the contingency time (Fig. 1d). Similarly, across experiments ($n=28$) there was a strong linear relationship between the contingency time and the time of maximal learning ($r=0.76$, $p<0.0001$, permutation test; slope of linear fit=0.92), and the mean learned trajectory (Fig. 1e, black trace) peaked within 2ms of the contingency time. These results suggest that the avian nervous system has temporally precise and accurate mechanisms both for encoding the timing of reinforcement and for directing learning.

Though maximal learning was accurately localized to the contingency time, the temporal extent of learning was broader than necessary to escape aversive reinforcement. In principle, successful escape could have been accomplished by changing only the part of the syllable monitored by the feedback delivery software, immediately surrounding the contingency time (Fig. 1e, gray dashed lines). Instead, learning spanned the entire targeted syllable (Fig. 1e, black trace), decaying gradually on either side of the contingency time (10–15% per 10ms). The observed structure of learning might reflect a fundamental time scale of premotor neural coding or result from peripheral musculoskeletal constraints. We considered an alternative possibility, that the structure of learning could be explained by the specific pattern of reinforced variation.

The structure of successful variation predicts learning

We evaluated the following simple model of learning: learn to produce the average of successful behavioral variants. For each experiment, the predictions of this model were calculated by computing the average of baseline fundamental frequency trajectories that avoid aversive reinforcement in an experimental simulation. We performed a simulation on baseline patterns of variation, as opposed to using the actual trials that the bird experienced during learning, to avoid the bias of including any of the actual structure of learning in the predictions. To simulate learning, we first normalized baseline fundamental frequency trajectories to yield residual trajectories expressed as percent deviations from the mean (Fig. 2a). Next, we simulated which of these trajectories would escape aversive reinforcement (Fig. 2b, red traces) given the contingency time and threshold for escaping reinforcement (see **Methods**). Our model predicts that the learned trajectory is the average of the trajectories that escape (i.e. the average of the red traces in Fig. 2b); we evaluated the model by comparing this predicted trajectory (Fig. 2c, red trace) to the actual learned trajectory (Fig. 2c, black trace).

The model accurately predicted learning. We compared the average prediction of learning for all experiments ($n=28$) to the average of actual learning for these experiments: the predicted trajectory of learning resided almost entirely within one standard error of the actual learned trajectory (Fig. 2d). Moreover, the coefficient of determination (r^2) between the predicted and actual learned trajectories was 0.94, meaning that 94% of the observed structure of learning was accounted for by the model.

These results demonstrate a correspondence between reinforced variation and learning on a millisecond time scale. One explanation is that the observed data reflect a fundamental time scale of song production and learning, implying that different patterns of learning are not possible. Such a fundamental time scale might arise from properties of premotor neural coding or by peripheral musculoskeletal constraints. In contrast, our model proposes that the pattern of learning in a given experiment does not reflect a fundamental limitation, but instead arises from the specific pattern of reinforced variation in that experiment. This model predicts that a more complex reinforcement contingency should elicit a distinct pattern of learning.

Dual contingency reinforcement yields predictable learning

To test the generality of our model, we performed a second set of experiments with a more complex reinforcement contingency. Whereas the first set of experiments required fundamental frequency to be above a threshold at one contingency time (single contingency experiments), this second set of experiments (dual contingency experiments) required fundamental frequency to be below a threshold at one contingency time (time A) and above a different threshold at a second contingency time 24ms later within the same syllable (time B, Fig. 3a). In both single and dual contingency experiments, a single aversive reinforcement signal was delivered at a specific time (Fig. 3a, time B). Thus if the nervous system only takes into account behavior at the time of reinforcement delivery, then learning should be similar in both experimental conditions. In contrast, if the nervous system learns the detailed trajectory of successful variation across the entire syllable, then learning in the dual contingency experiments should be downwards at the first contingency time and upwards at the second contingency time.

In dual contingency experiments, birds learned to shift fundamental frequency down at the beginning of the targeted syllable while shifting fundamental frequency up at the end of that syllable. We performed single and dual contingency experiments in a single syllable with feedback delivery at the same time (Fig. 3b–c). The learned change in the single contingency experiment was positive throughout the syllable, whereas the learned change in the dual contingency experiment was positive near contingency time B but negative earlier in the syllable. On average, learning in fundamental frequency in dual contingency experiments changed sign in this fashion 14.6ms before feedback delivery, whereas learning in single contingency experiments did not change sign across the entire targeted syllable (Fig. 3d). These results demonstrate that the learned trajectory in single contingency experiments does not represent a hard limit on the temporal specificity of learning; reinforcing a distinct subset of behavioral variants in dual contingency experiments results in more temporally specific learning.

The high temporal specificity of learning in dual contingency experiments was well predicted by the average of the fundamental frequency trajectories expected to escape aversive reinforcement (Fig. 3e). The coefficient of determination (r^2) between the average predicted and actual learning trajectories was 0.98. These results further support the model that the structure of successful variation predicts learning at a fine-grained level. Moreover,

results of the dual contingency experiments demonstrate that temporally sparse reinforcement is sufficient to elicit complex changes in behavioral trajectories.

The specific reinforcement history determines learning

The results of the single and dual contingency experiments suggest a correspondence between the average structure of variation that meets a fixed criterion and the average structure of learning when that variation is reinforced. To evaluate in a more detailed fashion whether the nervous system learns the average of successful behavioral trajectories, we tested whether learning in a given experiment could be predicted by the specific history of reinforced behavioral trajectories in that experiment. We performed a third set of experiments in which one aspect of the reinforcement contingency was fixed and instructive, but in which there was an additional stochastic element of the reinforcement contingency. Aversive reinforcement was applied not only to all fundamental frequency trajectories that failed to pass above a fixed threshold at the contingency time, as in single contingency experiments, but was also delivered to a random subset (50%) of the trajectories that did pass above the threshold (**see Methods**). These stochastic contingency experiments enabled us to test whether learning is better predicted by the trajectories passing above the threshold that escaped aversive reinforcement (random escapes) than by the trajectories passing above the threshold that received aversive reinforcement (random hits).

The learned trajectory was predicted more accurately by the average of random escapes than by the average of random hits (Fig. 4a). Prediction error was quantified as mean distance between predicted and actual learned frequency trajectories: $D = \text{mean}(\text{abs}(\text{actual} - \text{predicted}))$. Consistent with a model in which the specific history of reinforcement determines learning, prediction error was significantly smaller for the average of random escapes than for the average of random hits ($p=0.03$, Wilcoxon signed-rank test, $n=6$, Fig. 4b). Likewise, the average of random escapes was more correlated with the structure of actual learning ($p=0.03$, Wilcoxon signed-rank test, $n=6$).

Qualitatively, the predicted trajectories of learning in these experiments appeared to capture rapid timescale fluctuations in the actual trajectories of learning on a timescale of 5–10ms (Fig. 4a). Consistent with this observation, we found that predictions of learning were significantly poorer (Wilcoxon signed-rank test, $p<0.05$) when fundamental frequency fluctuations faster than ~12ms were filtered out of both the actual learned trajectory and the trajectories of the random escapes that constitute the prediction (Fig. 4c; **see Methods**). Across all experiments, prediction error increased as the timescale of the smoothing filter increased (Fig. 4d). Moreover, after smoothing, the prediction errors for the average of random hits and the average of random escapes were statistically indistinguishable. These results indicate that rapid fluctuations in fundamental frequency trajectories on a timescale of 12ms contribute to learning.

In principle, the prediction quality in single and dual contingency experiments could have resulted from a mixture of soft constraints that impose smoothness limitations on the structure of behavior and learning¹⁹, even in the absence of a specific relationship between successful behavioral performance and learning. In contrast, the results of the stochastic contingency experiments demonstrate that the nervous system records the specific history of

reinforcement and the structure of rapid timescale behavioral fluctuations and uses this information to guide learning.

Syllable-specific variation predicts learning

If the average structure of successful behavior indeed determines the structure of learning, then syllables with different patterns of fundamental frequency variation should exhibit different learning. Here, we found that some syllables exhibited slower fundamental frequency fluctuations than others (Fig. 5a–b) and we took advantage of this to test whether these natural differences in variation predict differences in learning.

Differences in the structure of fundamental frequency variation predicted differences in the structure of learning. For example, learning in syllables with faster fundamental frequency fluctuations (Fig. 5c, dashed green lines) was more temporally specific than learning in syllables with slower fluctuations (Fig. 5c, solid gray lines). For each experiment, we compared predictions of learning using the natural pattern of variation for the targeted syllable (“syllable-specific”) or using the patterns of variation for syllables in all other experiments (“general”). The prediction error was significantly lower when syllable-specific variability was used ($p=0.012$, Wilcoxon signed-rank test, $n=28$, Fig. 5d). These results show that syllable-specific differences in natural variation predict differences in learning, further demonstrating that the nervous system keeps track of the detailed structure of successful variation and uses it to guide learning.

The range of variation constrains learning

In a fourth set of experiments, we asked if the range of natural behavioral variation constrains learning from reinforcement. If the natural pattern of behavioral variation constrains learning, and all behavioral variants are paired with unfavorable outcomes, then the pattern of behavioral variation should remain constant and there should be no behavioral changes. Alternatively, the range of variation might expand to enable discovery of more favorable outcomes. Indeed, a previous study suggested that disruptive feedback persistently delivered to an individual syllable might drive increased variability in syllable structure²⁰.

To determine whether the natural range of variation constrains learning, we recorded baseline song (“baseline period”) and then delivered white noise to each syllable with fundamental frequency that fell within the baseline range of variation (“100% period”). Hence, during the 100% period, the bird could only receive an instructive learning signal (an escape) if he increased or changed the range of fundamental frequency variation. After the 100% period, which lasted a minimum of four days, an instructive period was used as a positive control for the bird's ability to learn. During the instructive period, white noise was delivered to the lowest (or highest) ninety percent of fundamental frequency performances.

Fundamental frequency performance was stable throughout the 100% period, whereas rapid and robust learning occurred during the instructive period (Fig. 6a, $n=5$). Furthermore, we found no evidence for an increase in trial-by-trial (Fig. 6b) or day-to-day (Fig. 6c) fundamental frequency variation during the 100% period. In summary, we saw no change in behavior when all variants of the behavior were unsuccessful. Our results demonstrate that

the natural pattern of behavioral variation constrains learning in the presence of a fixed reinforcement contingency.

Discussion

We found that the detailed structure of learning in response to reinforcement is predicted by the average of successful behavioral trajectories. In the first two sets of experiments, reinforcement delivery was contingent on the fundamental frequency trajectory of Bengalese finch song at one (Figs. 1 & 5) or two (Fig. 3) specific times in a 30–100ms long syllable. Although these experiments reinforced fundamental frequency trajectories with distinct temporal structure, in each case learning matched the average of successful trajectories. The third set of experiments revealed that learning matches the detailed history of successful behavioral trajectories even when the reinforcement contingency is partially stochastic (Fig. 4). The accuracy of these predictions at a millisecond time scale indicates that the nervous system has temporally precise mechanisms for tracking the trajectory of movements, encoding reinforcement, and directing learning. The fourth set of experiments demonstrated that learning does not occur when all behavioral trajectories are unsuccessful (Fig. 6). Together, these results indicate that the temporal structure of successful behavioral trajectories determines the pattern of learning. Thus, learning is structured by the temporal correlations present in natural behavioral variation.

This correlational learning rule implies that discrete reinforcement signals and simple averaging computations are sufficient to allow complex changes in behavioral trajectories. Moreover, this rule highlights the importance of behavioral variation as a substrate for learning. In birdsong, multiple time scales of variation are present in the temporal trajectories of fundamental frequency for individual syllables, allowing the same discrete reinforcement signal to elicit different patterns of learning in single (Figs. 1 & 5) and dual (Fig. 3) contingency experiments. From a broader perspective, our results indicate that the temporal structure of behavioral variation determines the shape of learned movement trajectories. Thus, fast time scales of variation allow temporally specific learning, whereas slow time scales of behavioral variation enable temporally extended behavioral changes in response to sparse reinforcement. Congruently, a transition towards faster time scales of variation as skill learning progresses would allow maintenance of gross behavioral parameters while facilitating subtle adjustments. Even in well-learned skills, however, diverse learning patterns could be highly adaptive, since skilled performance involves dynamics on a broad range of time scales²¹. For these reasons, the production of complex behavioral variation on multiple time scales may facilitate diverse patterns of learning in response to reinforcement.

Learning the average of successful performances has two potentially sub-optimal consequences. First, the capability for learning is limited by the natural pattern of variation. In action selection, there is evidence that animals can transcend this limitation by expanding the natural range of variation to explore for better outcomes²². In our experiments, however, learning did not occur if all behaviors received aversive reinforcement (e.g. Fig. 6). Manifestations of this limit have inspired incremental shaping techniques that modify human and animal behavior in an iterative manner, by applying differential reinforcement to

variation in the normal behavior, waiting for learning, and then applying reinforcement to variation in the learned behavior^{2,8,23–24}. The second sub-optimal consequence is that learning will reflect all aspects of behavior that are correlated with reinforcement, even in the absence of a causal relationship. This consequence could explain the development of superstitious behavior, such as the tendency of pigeons to learn whatever movements they happen to perform immediately prior to the stochastic and intermittent appearance of food reward⁴. Our single contingency experiments (Fig. 1) illustrate this principle on a rapid timescale by demonstrating that learning is not restricted to the contingency time (when fundamental frequency is causally related to outcome), but extends across the entire targeted syllable, in a manner accurately predicted by correlations in behavioral variation. Moreover, the results of our stochastic contingency experiments (Fig. 4) show that learning recapitulates the specific details of successful behavior, including the idiosyncrasies irrelevant to behavioral success.

Although our results demonstrate that the nervous system uses millisecond time scale information about behavioral variation for learning, they do not specify the sources of this variation or the mechanisms that convey information about successful variation to the nervous system. Variability is endemic to neural activity and behavioral performance, and a growing body of work has focused on the sources and consequences of such variation^{25–27}. In songbirds^{7,16,28–30} and mammals^{5–6,26,31–32}, trial-by-trial behavioral variation on a millisecond time scale is likely to result from both variation in the motor plan within the central nervous system and variation in the execution of that plan by the peripheral musculature. In principle, fine-grained variation from both central and peripheral sources could be reinforced. Operant learning has been elicited by delivering reinforcement contingent on neural variation in premotor regions^{33–34} (reflecting central variation) or somatosensory regions³⁵ (reflecting both central and peripheral variation). In songbirds, basal ganglia circuitry contributes to song variation and has been proposed as a primary source of variation for reinforcement learning^{17,36–45}. While our results do not identify the source of reinforced variation, the accuracy of our predictions suggests that much of fine-grained variation is susceptible to reinforcement. To learn from reinforcement, the nervous system must not only generate variable behavior but must also receive information about behavioral performance and outcome. Sensory feedback from auditory or proprioceptive channels might be sufficient to convey information about behavioral performance to the central nervous system. Alternatively, the central nervous system might keep a record, or efference copy, of the motor commands used to generate behavior. Although the relative contributions of these two mechanisms to learning remain unclear, our results (Fig. 4) implicate a mechanism that accurately encodes information about behavioral performance on a timescale of 10ms or faster.

Both single and dual contingency experiments revealed an exceptional capability for temporal specificity. Learning decayed significantly within tens of milliseconds before and after the contingency time. This temporal specificity is higher than for adaptive learning in eyeblink responses⁴⁶ and smooth pursuit eye movements⁴⁷. Moreover, in our dual contingency experiments, the direction of learning in the trajectories of fundamental frequency changed sign from positive to negative within 15ms of the second contingency time. This result demonstrates that independent control over adjacent time points in a

continuous behavior can be directed by sparse feedback signals acting on the complex correlational structure of behavioral variation. This temporal specificity presumably requires precise patterns of premotor neural activity and precisely controllable musculature, both of which have been reported in songbirds^{7,16,28–30,48}

Similarly precise mechanisms in humans may enable learning of the rapid acoustic modulations that confer meaning in spoken language and other precisely controlled movement trajectories^{49–50}.

Methods

Animal Care

Adult (> 100 day old) Bengalese finches were bred in our colony and housed with their parents until at least 60 days of age. During experiments, birds were isolated and housed individually in sound-attenuating chambers (Acoustic Systems) with food and water provided *ad libitum*. All song recordings were from undirected song (i.e. no female was present). All procedures were performed in accordance with established protocols approved by the University of California, San Francisco Institutional Animal Care and Use Committee.

Song acquisition and aversive reinforcement delivery

For single contingency experiments, song acquisition and feedback delivery were performed using previously described LabView software (EvTaf)⁸. EvTaf recognizes times in song when the spectral profile of song (measured in a 8ms window) matches a spectral template. We designed our spectral templates to uniquely identify a specific time (contingency time) in a specific syllable of song based on the spectral profile of song at that time and in the preceding 0–200ms (thus incorporating contextual information). This technique reduced the median standard deviation of the contingency time to 7.9ms; the remaining imprecision resulted from slight spectral differences between renditions of song and from the duration of the analysis window used to measure the spectral profile of song. Upon recognition, EvTaf recorded the time and calculated the fundamental frequency (FF) during the previous 8ms of song. If the FF met the escape criterion (i.e. above or below a threshold), then no disruptive feedback was delivered. Otherwise, a 50–80ms burst of white noise was delivered starting <1ms after the contingency time. The duration of white noise was constant for a given experiment. The escape criterion was set so that approximately 30% of syllables would escape disruptive feedback. In all experiments, birds adaptively changed the fundamental frequency of targeted syllables to avoid the white noise bursts. To allow quantification of FF during learning, a randomly interleaved 10% of songs were allocated as catch trials and did not receive any white noise.

For dual contingency experiments, EvTaf was modified to elicit a more complex learned trajectory. The modified software (EvTaf2) first recognized a contingency time and calculated FF. Then, instead of delivering differential feedback, EvTaf2 calculated the FF at a second contingency time 24ms later. EvTaf2 delivered a burst of white noise starting <1ms after the second contingency time unless FF at both contingency times met escape criteria.

To elicit temporally specific acoustic changes, these criteria were set in opposite directions (e.g. to escape white noise the FF had to be below FF_1 at contingency time 1 and above FF_2 at contingency time 2). Each of the escape criteria was set at approximately 60% so that the overall escape rate was similar to that in the single contingency experiments.

Experiments in which reinforcement delivery was partially stochastic were conducted similarly to single contingency experiments, but with several critical differences. When EvTaf detected that a syllable failed to meet the fixed escape criterion (i.e. above or below a set threshold), it delivered a burst of white noise on 100% of trials, as in single contingency experiments. However, when EvTaf detected a syllable that met the fixed escape criterion, it delivered a burst of white noise on a random 50% of trials (as opposed to 0% in single contingency experiments). To ensure a similar success rate as in single contingency experiments, the threshold was set lower, so that approximately 50% of syllables met the fixed criterion (and thus 25% of syllables successfully avoided white noise). To allow reliable quantification of fundamental frequency performance for each rendition of the targeted syllable, we used white noise stimuli that were low-pass filtered at 5000Hz and measured fundamental frequency trajectories from the third harmonic (at approximately 7000Hz) for both trials that escaped and received white noise. Since random differences between escapes and hits should diminish as sample size increases, we applied aversive reinforcement for a period of several hours (hundreds of trials) instead of several days (thousands of trials) as in previous experiments.

For 100% white noise experiments, the experimental design was the same as in single contingency experiments, except that the escape criterion was set >5 standard deviations from the mean, so that the bird would have to expand its range of FF variation in order to receive an instructive signal. Due to software limitations, a small percentage ($<2\%$) of syllables in each experiment failed to receive white noise, but these escapes were not correlated with FF of the targeted syllable and thus did not provide an instructive signal for FF adaptation. Of the five 100% white noise experiments, three included 10% catch trials and two did not include catch trials. In the experiments without catch trials, FF was computed during the earliest portion of the syllable before white noise playback. Results were statistically indistinguishable in experiments with and without catch trials, thus data from these experiments were combined for analysis. After at least four days, the threshold was lowered from >5 standard deviations from the mean to 1 standard deviation from the mean and maintained for at least three days to confirm that the bird was capable of learning when given an instructive signal.

Quantification of fundamental frequency variation and adaptation

To quantify the natural pattern of FF variation, we calculated residual FF trajectories as percent deviation from a stationary baseline (pre-feedback) mean. Stationary intervals typically lasted multiple hours and included hundreds of renditions of the targeted syllable.

The structure of FF adaptation was defined as the difference between the mean of baseline frequency trajectories and the mean of the frequency trajectories after learning had reached an asymptote. Asymptotic learning was defined as exceeding eighty percent of maximal adaptation in that experiment; analyses were robust to changes in this parameter.

Precise spectrograms (frequency vs. time representations) were calculated using a Gaussian-windowed short-time Fourier transform ($\sigma=1\text{ms}$) sampled at 8kHz. The duration of the Gaussian window was calibrated to the spectral sparseness of Bengalese finch harmonic stack notes using a previously developed technique¹⁸. Syllables were aligned by their onsets, based on amplitude threshold crossings. FF trajectories were computed by calculating the FF for each time bin in the spectrogram. In a given time bin, FF was determined by calculating the average FF for each harmonic weighted by the power in that harmonic. FF for each harmonic was calculated by parabolic interpolation.

To confirm the temporal resolution of this FF quantification algorithm, we generated artificial signals with a known FF trajectory and compared the estimated FF trajectories with the known signals. We used the Matlab (Mathworks) function 'vco' to generate these signals. Since we used a Gaussian filter with $\sigma=1\text{ms}$, we expected approximately millisecond-level resolution. First, we generated signals with uncorrelated FF trajectories (white noise) and found that the FF quantification algorithm did not generate spurious slow FF modulations. The FF trajectories generated from these signals exhibited a mean autocorrelational structure with a full-width at half-maximum of 1–2ms, consistent with millisecond resolution. Second, we generated signals with rapid FF transients and found that the FF quantification algorithm only slightly underestimated the magnitude of these transients. Only FF transients with duration less than 5ms were substantially underestimated (5ms was underestimated by 30%, 10ms was underestimated by <10%). In principle, song could have contained FF transients with duration less than 5ms, such that we were missing important structure in the data. To test this possibility, we estimated the duration of FF transients in our data set by calculating the power spectral density of all FF trajectories and correcting for the underestimation of rapid FF transients. Even after this correction, rapid FF transients of duration 1–5ms accounted for less than one percent of all FF variation. Together, these results indicate that the FF trajectories used in this study faithfully represent the acoustic structure of song.

Predicting learning from the structure of variation

We tested how well learning was predicted by the average of natural FF variants associated with relatively favorable outcomes (i.e. escape from disruptive feedback). For each experiment, we performed simulations to determine which baseline FF variants would escape disruptive feedback given the contingency time, adaptive direction, and threshold for escaping disruptive feedback (measured as the proportion of syllables that escaped, the "escape rate"). For single contingency experiments, the escape rate was approximately 30%. For consistency, and to determine that differences in adaptation could be predicted from differences in reinforced behavioral variation (as opposed to differences in escape rate), the same threshold (highest or lowest 30%, depending in the adaptive direction) was used for analysis of all single contingency experiments. Adjusting the simulated escape rate from 20–40% did not affect any of our conclusions. For dual contingency experiments, the escape rate at each contingency was approximately 60%. As with the single contingency experiments, a constant threshold (highest or lowest 60%, depending on the adaptive direction) was used for analysis, yet our conclusions were robust to varying the rate from 50–70%. Due to temporal imprecision in online recognition of the targeted syllable, the

contingency time varied slightly for different renditions of the same syllable in a given experiment (s.d. = 7.9 ms). To account for this imprecision, each experiment was simulated for the entire distribution of contingency times. The final prediction for a given experiment was the average of these simulations.

To determine whether rapid fluctuations in fundamental frequency contribute to the structure of learning, we smoothed fundamental frequency trajectories using a low-pass filter and measured the resulting change in prediction error. For the predictions of learning, the trajectories that randomly escaped were smoothed before the average was calculated. We low-pass filtered the signals in Matlab using a Butterworth filter with cutoff frequency defined as the frequency at which the magnitude response of the filter is the square root of 0.5. For this filter, a cutoff frequency of 83 Hz (12ms) reduces the power by >87% for fluctuations with a period of 10ms and reduces the power by <12% for fluctuations with a period of 15ms.

Statistical analysis

We used non-parametric tests for all statistical analysis. To assess the statistical significance of differences between paired data in two distinct conditions, we used the Wilcoxon signed-rank test. We used permutation tests to assess the statistical significance of linear relationships.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank P. Sabes, S. Sober, L. Frank, A. Doupe and S. Lisberger for discussion and comments on the manuscript and J. Wong and C. Brown for animal care. This work was supported by NIH R01 and P50 grants. E.C.T. was supported by an NIDCD NRSA postdoctoral fellowship and the Sloan-Swartz Foundation. J.D.C. and T.L.W. were supported by NSF graduate fellowships.

References

1. Thorndike, EL. Animal Intelligence. Macmillan; New York: 1911.
2. Skinner, BF. Science and Human Behavior. Macmillan; New York: 1953.
3. Balleine BW, Ostlund SB. Still at the choice-point: action selection and initiation in instrumental conditioning. *Ann. N. Y. Acad. Sci.* 2007; 1104:147–171. [PubMed: 17360797]
4. Skinner BF. Superstition in the pigeon. *J. Exp. Psychol.* 1948; 38:168–172. [PubMed: 18913665]
5. Schoppik D, Nagel KI, Lisberger SG. Cortical mechanisms of smooth eye movements revealed by dynamic covariations of neural and behavioral responses. *Neuron.* 2008; 58:248–260. [PubMed: 18439409]
6. Hira R, et al. Transcranial optogenetic stimulation for functional mapping of the motor cortex. *J. Neurosci. Methods.* 2009; 179:258–263. [PubMed: 19428535]
7. Fee MS, Kozhevnikov AA, Hahnloser RH. Neural mechanisms of vocal sequence generation in the songbird. *Ann. N. Y. Acad. Sci.* 2004; 1016:153–170. [PubMed: 15313774]
8. Tumer EC, Brainard MS. Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature.* 2007; 450:1240–1244. [PubMed: 18097411]
9. Tchernichovski O, Mitra PP, Lints T, Nottebohm F. Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science.* 2001; 291:2564–2569. [PubMed: 11283361]

10. Franz M, Goller F. Respiratory units of motor production and song imitation in the zebra finch. *J. Neurobiol.* 2002; 51:129–141. [PubMed: 11932954]
11. Mendez JM, Dall'asén AG, Cooper BG, Goller F. Acquisition of an acoustic template leads to refinement of song motor gestures. *J. Neurophysiol.* 2010; 104:984–993. [PubMed: 20554848]
12. Ashmore RC, Wild JM, Schmidt MF. Brainstem and forebrain contributions to the generation of learned motor behaviors for song. *J. Neurosci.* 2005; 25:8543–8554. [PubMed: 16162936]
13. Glaze CM, Troyer TW. Temporal structure in zebra finch song: implications for motor coding. *J. Neurosci.* 2006; 26:991–1005. [PubMed: 16421319]
14. Glaze CM, Troyer TW. Behavioral measurements of a temporally precise motor code for birdsong. *J. Neurosci.* 2007; 27:7631–7639. [PubMed: 17634357]
15. Tchernichovski O, Lints TJ, Deregnacourt S, Cimenser A, Mitra PP. Studying the song development process: rationale and methods. *Ann. N. Y. Acad. Sci.* 2004; 1016:348–363. [PubMed: 15313784]
16. Chi Z, Margoliash D. Temporal precision and temporal drift in brain and behavior of zebra finch song. *Neuron.* 2001; 32:899–910. [PubMed: 11738034]
17. Andalman AS, Fee MS. A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc. Natl. Acad. Sci. U. S. A.* 2009; 106:12518–12523. [PubMed: 19597157]
18. Gardner TJ, Magnasco MO. Sparse time-frequency representations. *Proc. Natl. Acad. Sci. U. S. A.* 2006; 103:6094–6099. [PubMed: 16601097]
19. Todorov E. Optimality principles in sensorimotor control. *Nat. Neurosci.* 2004; 7:907–915. [PubMed: 15332089]
20. Leonardo A, Konishi M. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature.* 1999; 399:466–470. [PubMed: 10365958]
21. Mauk MD, Buonomano DV. The neural basis of temporal processing. *Annu. Rev. Neurosci.* 2004; 27:307–340. [PubMed: 15217335]
22. Cohen JD, McClure SM, Yu AJ. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2007; 362:933–942. [PubMed: 17395573]
23. Staddon, JER. *Adaptive behavior and Learning.* Cambridge University Press; Cambridge, U.K.: 1983.
24. Catania, AC. *Learning.* 2nd edition. Prentice Hall; Englewood Cliffs, N.J.: 1984.
25. Churchland MM, Afshar A, Shenoy KV. A central source of movement variability. *Neuron.* 2006; 52:1085–1096. [PubMed: 17178410]
26. Faisal AA, Selen LP, Wolpert DM. Noise in the nervous system. *Nat. Rev. Neurosci.* 2008; 9:292–303. [PubMed: 18319728]
27. Todorov E, Jordan MI. Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* 2002; 5:1226–1235. [PubMed: 12404008]
28. Elemans CP, Spierts IL, Müller UK, Van Leeuwen JL, Goller F. Bird song: superfast muscles control dove's trill. *Nature.* 2004; 431:146. [PubMed: 15356620]
29. Elemans CP, Mead AF, Rome LC, Goller F. Superfast vocal muscles control song production in songbirds. *PloS One.* 2008; 3:e2581. [PubMed: 18612467]
30. Sober SJ, Wohlgenuth MJ, Brainard MS. Central contributions to acoustic variation in birdsong. *J. Neurosci.* 2008; 28:10370–10379. [PubMed: 18842896]
31. van beers RJ. The sources of variability in saccadic eye movements. *J. Neurosci.* 2007; 27:8757–8770. [PubMed: 17699658]
32. van beers RJ, Haggard P, Wolpert DM. The role of execution noise in movement variability. *J. Neurophysiol.* 2004; 91:1050–1063. [PubMed: 14561687]
33. Fetz EE, Finocchio DV. Operant conditioning of specific patterns of neural and muscular activity. *Science.* 1971; 174:431–435. [PubMed: 5000088]
34. Rosenfeld JP, Hetzler BE. Operant-controlled evoked responses: discrimination of conditioned and normally occurring components. *Science.* 1973; 181:767–770. [PubMed: 4724935]
35. Rosenfeld JP, Fox SS. Operant control of a brain potential evoked by a behavior. *Physiol. Behav.* 1971; 7:489–493. [PubMed: 5131207]

36. Bottjer SW, Miesner EA, Arnold AP. Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science*. 1984; 224:901–903. [PubMed: 6719123]
37. Scharff C, Nottebohm F. A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *J. Neurosci*. 1991; 11:2896–2913. [PubMed: 1880555]
38. Jarvis ED, Scharff C, Grossman MR, Ramos JA, Nottebohm F. For whom the bird sings: context-dependent gene expression. *Neuron*. 1998; 21:775–788. [PubMed: 9808464]
39. Hessler NA, Doupe AJ. Social context modulates singing-related neural activity in the songbird forebrain. *Nat. Neurosci*. 1999; 2:209–211. [PubMed: 10195211]
40. Doya, K.; Sejnowski, TJ. A computational model of avian song learning. In: Gazzaniga, MS., editor. *The New Cognitive Neurosciences*. MIT Press; Cambridge, Massachusetts: 2000. p. 469–482.
41. Kao MH, Doupe AJ, Brainard MS. Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature*. 2005; 433:638–643. [PubMed: 15703748]
42. Olveczky BP, Andalman AS, Fee MS. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol*. 2005; 3:e153. [PubMed: 15826219]
43. Fiete IR, Fee MS, Seung HS. Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *J. Neurophysiol*. 2007; 98:2038–2057. [PubMed: 17652414]
44. Aronov D, Andalman AS, Fee MS. A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science*. 2008; 320:630–634. [PubMed: 18451295]
45. Hampton CM, Sakata JT, Brainard MS. An avian basal ganglia-forebrain circuit contributes differently to syllable versus sequence variability of adult Bengalese finch song. *J. Neurophysiol*. 2009; 101:3235–3245. [PubMed: 19357331]
46. Mauk MD, Ruiz BP. Learning-dependent timing of Pavlovian eyelid responses: differential conditioning using multiple interstimulus intervals. *Behav. Neurosci*. 1992; 106:666–681. [PubMed: 1503659]
47. Medina JF, Carey MR, Lisberger SG. The representation of time for motor learning. *Neuron*. 2005; 45:157–167. [PubMed: 15629710]
48. Leonardo A, Fee MS. Ensemble coding of vocal control in birdsong. *J. Neurosci*. 2005; 25:652–661. [PubMed: 15659602]
49. House, D. *Tonal perception in speech*. Lund University Press; Lund, Sweden: 1990.
50. Hermes, DJ. Stylization of pitch contours. In: Sudhoff, S., editor. *Methods in Empirical Prosody Research*. Walter de Gruyter; Berlin: 2006. p. 29–62.

Figure 1. The microstructure of learning from precisely timed reinforcement

a. For each rendition of targeted syllable “A” (e.g. A1 and A2), a burst of white noise (aversive reinforcement) was delivered with minimal delay ($<1\text{ms}$) unless fundamental frequency was higher than a threshold value at the contingency time (blue arrows). Thus, A1 escapes while A2 receives aversive reinforcement. The duration of white noise (50–80ms) was constant in a given experiment. **b.** Adaptive change in fundamental frequency (FF) following contingent aversive reinforcement. Post-reinforcement FF was measured after learning has reached asymptote (see Methods). **c.** Syllable A before (Pre) and after (Post) reinforcement based on fundamental frequency at the contingency time shown by blue arrowheads. An enlarged view of the first harmonic shows temporally precise representations of mean FF (blue traces) and reveals an upward change in FF around the contingency time. The temporal structure of learning was quantified by dividing mean FF performance after reinforcement by mean FF performance before reinforcement and normalizing by maximal change in the adaptive direction. **d.** Learning for three experiments in one syllable with distinct contingency times. **e.** Mean learning (thick black line) and learning for individual experiments (thin colored lines) aligned to the contingency time (blue arrowhead) for each experiment ($n=28$). In principle, consistent escape of aversive feedback could have been achieved by changing fundamental frequency only within the range of contingency times (between gray dashed lines), but actual learning extended well beyond this temporal window.

Figure 2. The microstructure of successful variation predicts learning

a. To describe the natural pattern of FF variation, we calculated temporally precise representations of FF for baseline performances of the targeted syllable. The left panel shows the mean spectrogram of this syllable at baseline. In the middle panel, temporally precise FF representations (black traces) are overlaid on spectrograms of the first harmonic for two baseline performances of the syllable. The right panel depicts temporally precise FF representations for 50 baseline performances, expressed as percent deviations from the mean. **b.** We predicted learning from the baseline structure of FF variation by computing the average of the baseline FF variants that avoid aversive reinforcement in a simulation of the experiment. Simulations included information about the contingency time (blue arrowhead) and threshold for avoiding aversive reinforcement (upper tip of blue arrowhead) for a given experiment (see Methods). In this example, the simulation indicated that the red trajectories would avoid aversive reinforcement and the gray trajectories would receive aversive reinforcement. **c.** Predicted learning (red) compared to actual learning (black) in this example experiment. **d.** Average predicted (red) and actual (black) learning trajectories across all experiments (n=28). Gray shading denotes $\pm$ s.e.m. for actual learning. All traces are aligned to the contingency time.

Figure 3. The microstructure of successful variation predicts learning with a complex reinforcement contingency

a. We compared learned FF trajectories for two different tasks. In the “single contingency” task described in Fig. 1, white noise was delivered at time B unless FF at time B was above a threshold value. In the “dual contingency” task, white noise was delivered at time B unless FF at time B was above a threshold value and FF at time A was below a threshold value. **b.** Calculating predictions of learning for single contingency (top) and dual contingency (bottom) experiments in one syllable. As before, learning in a given experiment was predicted by computing the average of baseline FF variants that avoid aversive reinforcement in a simulation of the experiment (red traces). **c.** Actual (black) and predicted (red) learning for single and dual contingency experiments for this syllable. **d.** Comparison of average dual contingency (black, $n=8$) and single contingency (green, $n=28$) learning, normalized to adaptation at contingency time B. The downward arrowheads indicate the magnitude of learning for individual experiments at 12.5 ms before the contingency time. Learned trajectories for single and dual contingency experiments were non-overlapping at this time and all earlier times in the syllable. **e.** Average predicted (red) and actual (black) learned trajectories across all dual contingency experiments (mean $\pm$ 1 s.e.m).

Figure 4. The specific history of reinforcement determines the structure of learning

a. Comparison of predicted and actual learning for a stochastic contingency experiment in which aversive reinforcement was not only delivered to fundamental frequency trajectories that were below a threshold at a single contingency time (blue arrowhead) but was also delivered to a random 50% of fundamental frequency trajectories that were above the threshold. Actual learning (black) is compared to the average of the trajectories that exceeded the threshold and randomly escaped aversive reinforcement (“random escapes”, left panel, blue trace) or the average of the trajectories that exceeded the threshold yet randomly received aversive reinforcement (“random hits”, right panel, red trace). Gray shading represents prediction error. **b.** In all experiments, the prediction error was greater for the average of random hits than for the average of random escapes ($n=6$, $p=0.03$, Wilcoxon ranked-sign test). **c.** To test whether rapid fluctuations in fundamental frequency contribute to prediction accuracy, we took the data shown in the left panel of **a** and smoothed both the actual learned trajectory and the individual trajectories that contributed to the prediction, thus attenuating fluctuations more rapid than 12ms (see Methods). **d.** For the six experiments illustrated in **b**, prediction error increased as we increased the low-pass cutoff frequency with which the raw data were smoothed. Filtering out fluctuations faster than 12ms (arrow) resulted in a significant increase in prediction error (asterisks, Wilcoxon signed-rank test) relative to unfiltered data (blue dashed line).

Figure 5. Inter-syllable differences in the structure of variation predict differences in learning
a. Natural pattern of fundamental frequency variation for two syllables, expressed as residuals from the mean. Each panel depicts the fundamental frequency performance on 30 consecutive renditions of the syllable during baseline song. The variability for syllable B (green) appears to exhibit faster temporal fluctuations than for syllable A (gray). **b.** The time scale of variability for syllables A (gray) and B (green, dashed). Coefficient of determination (r^2) traces depict the extent to which a deviation from the mean at a specific time in the syllable determines the deviations from the mean at surrounding times in the syllable. A more rapid decay in the coefficient of determination indicates a more rapid time scale of variability (e.g. syllable B relative to syllable A). **c.** Actual learning in syllables A and B, compared with predictions of learning calculated as before. **d.** We compared the actual learning for a syllable with predictions based on FF variation for that syllable (“syllable-specific” variation, as before) or predictions based on FF variation for syllables targeted in all other experiments (“general” variation). Prediction error was quantified as mean distance from the actual structure of learning. Predictions using syllable-specific variation had significantly less error than predictions using general variation (Wilcoxon signed-rank test; $p=0.005$, $n=28$; horizontal lines denote means).

Figure 6. The range of variation constrains learning

a. Summary data for experiments ($n=5$) that consisted of a baseline period, followed by a 100% aversive reinforcement period, followed by an instructive aversive reinforcement period. Circles and vertical lines indicate mean $\pm$ s.e.m. of FF for an entire day. The baseline period was used to characterize the natural range of within-day and between-day FF variation. In the 100% period, all syllables received white noise. In the instructive period, all syllables with high FF were allowed to escape white noise. Days 7 and 11 correspond to the final day in the 100% and instructive period, respectively. **b.** Comparison of within-day FF variation. The CV of FF was calculated for each of the baseline and 100% days shown in A. **c.** Comparison of between-days FF variation. Change in mean FF between days was computed for each of the pairs of days shown. In **a–c**, vertical lines denote $\pm$ s.e.m.