Goal-directed vocal planning in a songbird

Authors

Anja T. Zai[1,2]*, Anna E. Stepien[1,2], Nicolas Giret[3], Richard H.R. Hahnloser[1,2]

Affiliations

[1]Institute of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland.

[2]Neuroscience Center Zurich (ZNZ), University of Zurich and ETH Zurich, Zurich, Switzerland.

[3]Institut des Neurosciences Paris Saclay, UMR 9197 CNRS, Université Paris Saclay.

* corresponding author: zaia@ethz.ch

Abstract

Songbirds' vocal mastery is impressive, but to what extent is it a result of practice? Can they, based on experienced mismatch with a known target, plan the necessary changes to recover the target in a practice-free manner without intermittently singing? In adult zebra finches, we drive the pitch of a song syllable away from its stable (baseline) variant acquired from a tutor, then we withdraw reinforcement and subsequently deprive them of singing experience by muting or deafening. In this deprived state, birds do not recover their baseline song. However, they revert their songs towards the target by about one standard deviation of their recent practice, provided the sensory feedback during the latter signaled a pitch mismatch with the target. Thus, targeted vocal plasticity does not require immediate sensory experience, showing that zebra finches are capable of goal-directed vocal planning.

MAIN TEXT

Introduction

Speech planning is an important part of human communication and the inability to plan speech is manifest in disorders such as apraxia. But to what extent is targeted vocal planning an entirely human ability? Many animals are capable of volitional control of vocalizations (1,2), but are they also capable of planning to selectively adapt their vocalizations towards a target, such as when striving to reduce the pitch mismatch of a note in a song? Target-specific vocal planning is a cognitive ability that requires extracting or recalling a sensory target and forming or selecting the required motor actions to reach the target. Such planning can be covert or overt. Evidence for covert planning is manifest when a targeted motor change is executed without intermittent practice (3), e.g., when we instantly imitate a word upon first hearing. Overt planning, by contrast, includes practice, but without access to the sensory experience from which target mismatch could be computed, e.g., when we practice a piano piece by tapping on a table.

The vocal planning abilities in animals and their dependence on sensory experience remain poorly explored. Motor learning has been mostly studied in tasks where a skilled behavioral response must be produced on the spot, such as when a visual target must be hit by a saccade or by an arm reaching movement (4–7). In this context, motor planning has been shown to enhance motor flexibility, as it allows separation of motor memories when

there are conflicting perturbations (8). However, for developmental behaviors such as speech or birdsong that rely on hearing a target early in life (9,10), the roles of practice and of sensory feedback for flexible vocal control and for target-directed adaptation are unknown.

Recovery of a once-learned vocal skill could be instantaneous (covert), or it might require practice (overt). In support of the former, many motor memories are long-lasting (11), e.g., we can recall the happy-birthday song for years without practice. Some memories are even hard to get rid of such as accents in a foreign language. By contrast, practice-dependent, but feedback-independent recovery is argued for by arm reaching movements during use-dependent forgetting: following adaptation to biasing visual feedback, arm movements recover when the bias is either removed or the visual error is artificially clamped to zero (5,6). One explanation put forward is that motor adaptation is volatile and has forgetting built-in (7,12), leading to practice-dependent reappearance of the original motor program even without informative feedback (12). Given these possibilities, we set out to probe songbirds' skills of recovering their developmental song target when deprived of either singing practice (to probe covert planning) or of sensory feedback (to probe overt planning).

Adult vocal performances in songbirds can be altered by applying external reinforcers such as white-noise stimuli (13,14). When the reinforcer is withdrawn, birds recover their original song within hundreds of song attempts (13,15–17). We argued that these attempts may be unnecessary and birds could recover their original performance by recalling either 1) the original motor program (18–20), or 2) its sensory representation (21–23) plus the mapping required for translating that into the original program (15,16) (Fig. 1A). These options might not need sensory feedback, which is argued for by birds' large perceptual song memory capacity (24). That is, birds' song practice may be mainly expression of deliberate playfulness (25), conferring the skill of vocal flexibility rather than serving to reach a target, evidenced by young birds that explore vocal spaces close to orthogonal to the song-learning direction (26) and that are already surprisingly capable of adult-like singing when appropriately stimulated (27).

Results

To test whether birds can covertly recover a song syllable without practice, we first reinforced the pitch of a song syllable away from baseline and then we suppressed birds' singing capacity for a few days by muting their vocal output. We then unmuted birds and tested whether the song has covertly reverted back to the original target. We used syllable pitch as the targeted song feature, because we found that birds did not reliably recover syllable duration in experiments in which we induced them to shorten or lengthen syllable duration (Fig. S1).

We first drove pitch away from baseline by at least one standard deviation using a white-noise (WN) stimulus delivered whenever the pitch within a 16-ms time window locked to the targeted syllable was above or below a manually set threshold (Fig. 1B, C, see Methods). We muted these WNm (white-noise reinforced and muted) birds by implanting a bypass cannula into the abdominal air sac (see Methods). While muted, air is leaking from the abdominal air sac and as a result, sub-syringeal air pressure does not build up to exceed the threshold level required for the self-sustained syringeal oscillations (28) that

underlie singing. Physical absence of such oscillations essentially strips muted birds from all pitch experience. In some cases, the bypass cannula got clogged during the muted period and birds were spontaneously unmuted, allowing them to produce a few songs before we reopened the cannula (Fig. 1C-G).

We quantified recovery in terms of normalized residual pitch (NRP) to discount for differences in the amount of initial pitch shift where $NRP = 0\%$ corresponds to complete recovery and $NRP = 100\%$ corresponds pitch values before withdrawal of reinforcement (R) and thus no recovery. After spending 5.1 ± 1.6 days (range $3 - 8$ d, $N = 8$) in the muted state and upon unmuting, WNm birds displayed an average NRP of 89%, which was far from baseline ($p = 6.2 \cdot 10^{-8}$, tstat $= -23.6$, $N = 8$ birds, two-sided t-test of H0: $NRP = 0\%$, songs analyzed in 2 h time window – early (E), see methods, Fig. 1), suggesting that in the muted state, birds are unable to recover their pre-reinforced songs. The average NRP in WNm birds was comparable to that of unmanipulated control (WNC) birds within the first 2 h after withdrawal of the reinforcer (average $NRP = 91\%$, $p = 3.7 \cdot 10^{-11}$, tstat $= 14.8$, two-sided t-test for $NRP = 0\%$, $N = 18$ WNC birds). Indeed, during 5 days without song practice, birds recovered no more pitch distance than birds normally do within the first 2 h of release from reinforcement ($p = 0.82$, tstat $= -0.23$, $N = 8$ WNm and $N = 18$ WNC birds, two-sided t-test). In WNm birds, there was no correlation between the NRP in the early window and the time since the muting surgery (correlation coefficient $= 0.26$, $p = 0.53$), suggesting that the lack of pitch recovery while muted was not due to a lingering burden of the muting surgery. These findings did not sensitively depend on the size of the analysis window — we also tested windows of 4 and of 24 h.

Subsequently, after 4 days of unmuted singing experience (roughly 9 days after withdrawal of WN), WNm birds displayed an average NRP of 30%, which was significantly different from the average NRP within the first 2 h after unmuting ($p = 3 \cdot 10^{-4}$, tstat $= 4.83$, $N = 8$ birds, two-tailed t-test early (E) vs. late (L) time window) but still significantly different from zero ($p = 0.04$, tstat $= 2.59$, $N = 8$ birds, two-tailed t-test, late (L) time window). The amount of recovery was neither correlated with the number of renditions sung between early and late windows (correlation coefficient $= 0.03$, $p = 0.95$), nor with the duration the birds were muted (correlation coefficient $= -0.50$, $p = 0.20$), nor with the time since they last sung the target song before reinforcement (correlation coefficient $= -0.43$, $p = 0.29$), suggesting the limiting factor for recovery was neither the amount of song practice nor the recovery time from the muting surgery (although for the latter there was a trend). Overall, these findings rule out covert planning in muted birds and suggest that motor practice is necessary for recovery of baseline song.

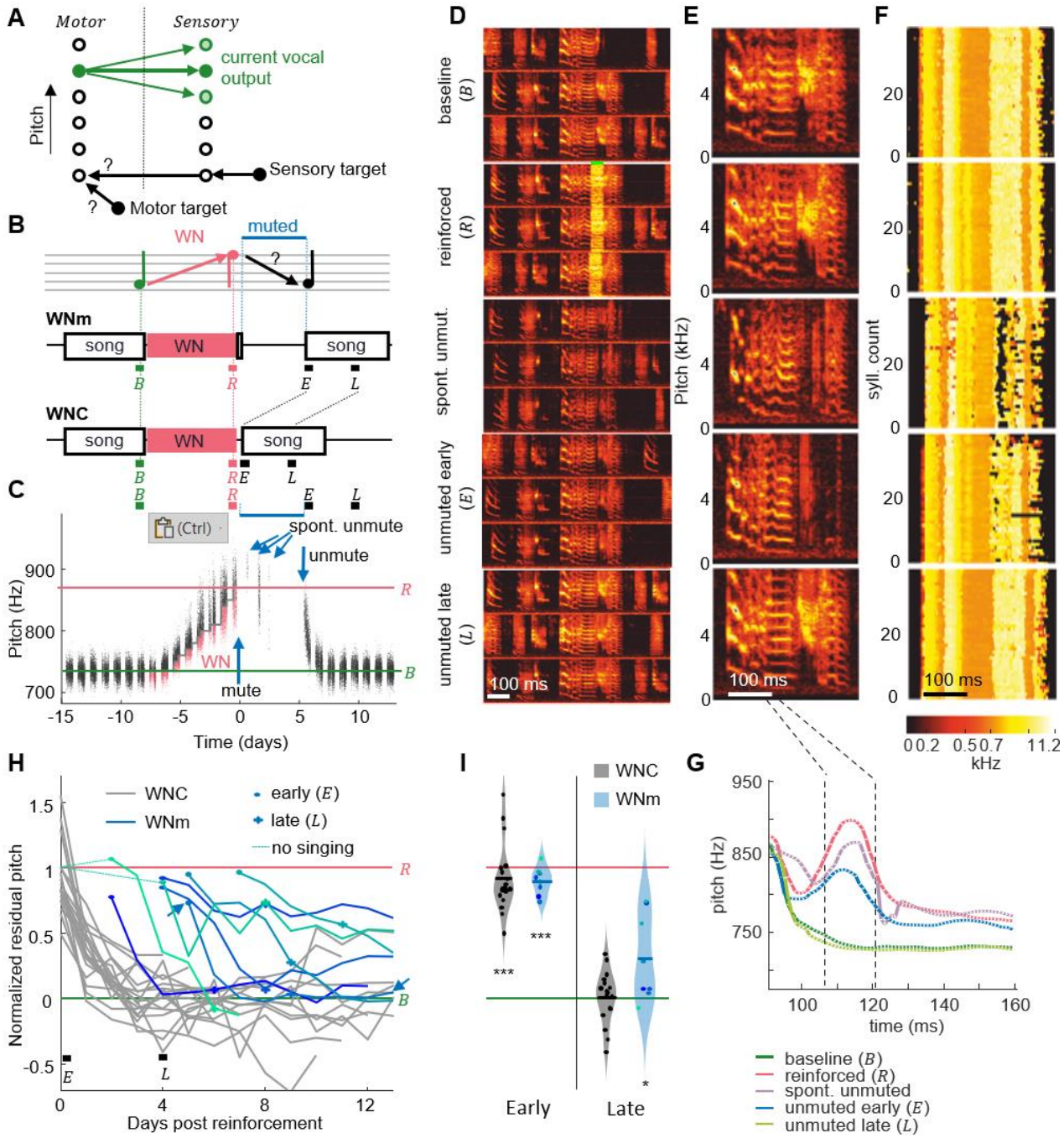

Fig. 1. Recovery of pitch target requires practice. (**A**) Two hypotheses on birds' ability to recover a song target away from their current vocal output (green circles, motor states on the left, sensory states on the right, shading represents probabilities): Either they could recall the motor target and reactivate it without practice, or they could recall a sensory target plus the neural mapping (black arrows) required to transform it into a motor state. (**B**) WNm birds were first pitch-reinforced using white noise (WN), then muted, and subsequently unmuted. WN was delivered when the pitch of the target syllable was either below (as exemplified here) or above a threshold. Pitch recovery from the reinforced (R) state towards the baseline (B) target is evaluated in early (E, no practice) and late (L, with practice) analysis windows (all windows are time-aligned to the first 2 h of songs after withdrawal of reinforcement, E) and compared to recovery in unmuted control

birds (WNC). (**C**) Syllable pitches (dots, red=reinforced syllables) of an example bird that while muted recovered only about 27% of pitch difference to baseline despite three spontaneous unmuting events (arrows). (**D**) Same bird, spectrograms of example song motifs from 5 epochs: during baseline (B), reinforcement (R) with WN (green bar), spontaneous unmuting (spont. unmut.), and during permanent unmuting (early – E and late - L). (**E**) Example syllables from same 5 epochs. (**F**) Stack plot of pitch traces (pitch indicated by color, see color scale) of the first 40 targeted syllables in each epoch ('reinforced': only traces without WN are shown). (**G**) Average pitch traces from F), revealing a pitch increase during the pitch-measurement window (dashed black lines) and pitch recovery late after unmuting. (**H**) WNm birds (blue lines, $N = 8$) showed a normalized residual pitch (NRP) far from zero several days after reinforcement (circles indicate unmuting events, arrow shows bird from C) unlike WNC birds (gray lines, $N = 18$). Thin dashed lines indicate the two initial birds that were not given reinforcement-free singing experience before muting (see Methods). (**I**) Violin plots of same data restricted to early and late analysis windows (***$p < 0.001$, *$p < 0.05$, two-tailed t-test of $NRP = 0$).

Next, we tested whether motor experience but not sensory experience is necessary for overt recovery, similar to arm reaching movements that can be restored without guiding feedback (5,29). In a second group of birds, we provided slightly more singing experience (Fig. 2). Instead of muting, WNd birds were deafened through bilateral cochlea removal immediately after the end of WN reinforcement. This latter manipulation does not suppress the act of singing as does muting, but it eliminates auditory feedback from singing. Deaf birds could gain access to some pitch information via somatosensory stretch and vibration receptors and/or air pressure sensing (30). Our aim was to test whether such putative pitch correlates are sufficient for recovery of baseline pitch (Fig. 2A). However, in the deaf state, WNd birds did not recover baseline pitch even after 4 days of song practice: on the 5th day (late, L) after deafening, their average NRP was still 50%, which was different from zero ($p = 0.03$, tstat $= 2.73$, two-tailed t-test of H0: $NRP = 0\%$, $N = 10$, Fig. 2D) and significantly larger than the average NRP of WNC birds on the 5th day since withdrawal of reinforcement (difference in $NRP = 49\%$, $p = 0.003$, tstat $= 3.34$, $df = 26$, $N = 10$ WNd and $N = 18$ WNC birds, two-tailed t-test).

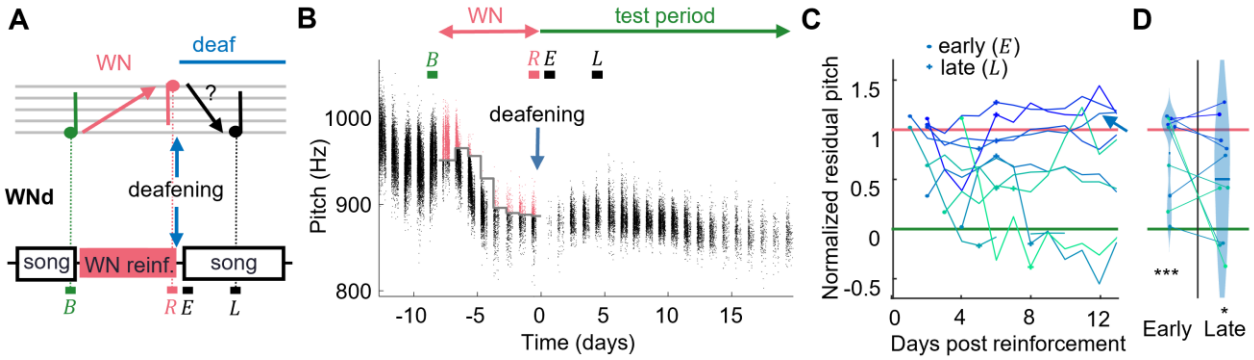

Fig. 2. Recovery of pitch target is impaired after deafening. (**A**) WNd birds were first pitch-reinforced using white noise (WN) and then deafened by bilateral cochlea removal. Analysis windows (letters) as in Fig. 1. (**B**) Syllable pitches (dots, red=reinforced syllables) of example WNd bird that shifted pitch down by $d' = -2.7$ during WN reinforcement and subsequently did not recover baseline pitch

during the test period. (**C**) WNd birds ($N = 10$) do not recover baseline pitch without auditory feedback (circles=early window after deafening events, cross=late). (**D**) Violin plots of same data restricted to early and late analysis windows, lines connect individual birds (***$p < 0.001$, *$p < 0.05$, two-tailed t-test of $NRP = 0$).

We speculated that the lack of pitch recovery in WNd birds could be attributable to the sudden deafening experience, which might be too overwhelming to uphold the plan to recover the original pitch target. WN deaf birds did not sing for an average of 2.3 ± 1.1 days (range 1 to 4 days) after the deafening surgery, which is a strong indication of an acute stressor (31). We thus inspected a third group of birds (dLO, Fig. 3) taken from (32) that learned to shift pitch while deaf and that underwent no invasive treatment between the pitch reinforcing experience and the test period of song recovery.

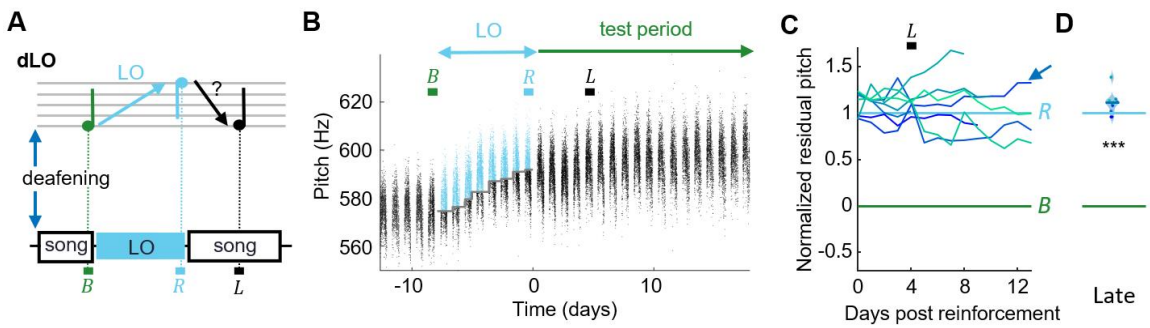

Fig. 3. Deaf birds do not recover pitch target after light-induced mismatch. (**A**) dLO birds were first deafened and then pitch-reinforced using a brief light-off (LO) stimulus. Analysis windows (letters) as in Fig. 1. (**B**) Syllable pitches (dots, blue=LO-reinforced syllables) of example dLO bird that shifted pitch up by $d' = 3.5$ within a week, but showed no signs of pitch recovery during the test period. (**C**) dLO birds ($N = 8$) do not recover baseline pitch without auditory feedback. (**D**) Violin plots of same data restricted to the late analysis window (***$p < 0.001$, two-tailed t-test of $NRP = 0$).

dLO birds were first deafened, and after they produced stable baseline song for several days, their target syllable pitch was reinforced using pitch-contingent light-off (LO) stimuli, during which the light in the sound recording chamber was briefly turned off upon high- or low-pitch syllable renditions (33). dLO birds displayed an average NRP of 112% on the 5th day since release from LO, which was significantly different from zero ($p = 3.7 \cdot 10^{-8}$, tstat $= 25.4$, $N = 8$ birds, two-tailed t-test of H0: NRP=0) and was larger than the NRP in WNC birds on the 5th day since release ($p = 1.3 \cdot 10^{-13}$, tstat $= 14.9$, $df = 24$, $N = 8$ dLO and $N = 18$ WNC birds, two-sided t-test). Thus, dLO birds were unable to recover baseline pitch, suggesting that song recovery requires undiminished sensory experience, which includes auditory feedback.

Deaf birds' decreased singing rate could not explain their lack of pitch recovery. Deaf birds sang less during the first 2 hours since release of reinforcement (early) than control birds: 87 ± 59 motif renditions for WNd and 410 ± 330 renditions for dLO compared to 616 ± 272 renditions for WNC birds. Also, WNd birds sang only 4300 ± 2300 motif renditions between the early and late period compared to the average of 11000 ± 3400 renditions that hearing WNC birds produced in the same time period. However, despite these differences, when we inspected WNd birds' behavior 9 days after the early window, when they sung on average 12000 ± 6000 renditions, their NRP was still

significantly different from zero ($NRP = 0.37$, $p = 0.007$, tstat $= 3.47$, $df = 9$). Thus, even after producing more practice songs than control birds, deaf birds did not recover baseline pitch and so the number of songs alone cannot explain why deaf birds do not fully recover pitch. We conclude that auditory experience seems to be necessary to recover song.

That song practice and sensory experience are required for full recovery of song does not imply that without experience, birds are incapable of making any targeted changes to their songs at all. We therefore inspected birds' fine-grained vocal output and whether they changed their song in the direction of baseline when deprived of sensory experience. We hypothesized (Fig. 4A) that when birds experience a target mismatch during reinforcement (i.e., they hear that their song deviates from the target), they plan to recover the pitch target, and a portion of this plan they can execute without feedback. If, by contrast, they have no mismatch experience before deafening, they will make no corresponding plan. Hence, we predicted that WNd birds that experienced a pitch mismatch during reinforcement and before deafening would slightly revert their song towards baseline even in the absence of auditory feedback. By contrast, dLO birds that did not experience a mismatch because they did not hear their song while it was reinforced, would not revert towards the target (Fig. 4A).

Indeed, WNd birds changed their pitch significantly towards baseline already in the first 2 h of their singing since release from reinforcement (relative to the pitch from the last 2 h during reinforcement). We quantified local pitch changes in terms of the d' sensitivity of signal detection theory (which is independent of shift magnitude) and found $d' = -0.60$ ($p = 0.03$, tstat $= -2.19$, $df = 9$, $N = 10$ WNd birds, one-sided t-test of H0: $d' = 0$). A significant reversion towards pitch baseline was still evident after 4 days of practice ($d' = -1.27$, $p = 0.02$, tstat $= -2.35$, $N = 10$ WNd birds, one-sided t-test, Fig. 4B, D, F), showing that pitch reversion in deaf birds is persistent. Because the average pitch shift in WNd birds was on the order of one standard deviation ($d' \simeq 1$), we conclude that without auditory experience, birds are able to perform target-directed pitch shifts of about the same magnitude as their current exploratory range (i.e., the denominator of the d' measure).

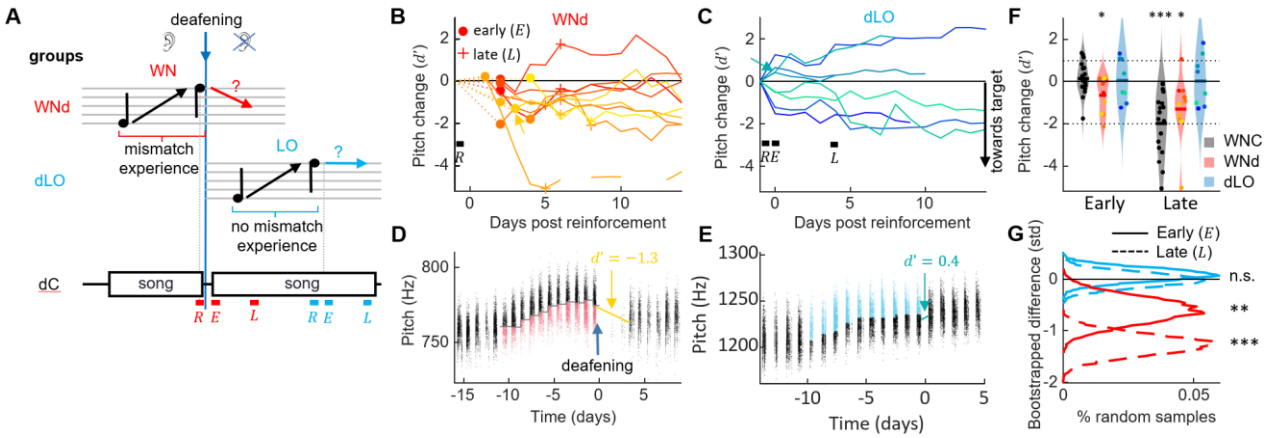

Fig. 4. Target mismatch experience is necessary for revertive pitch changes. (**A**) WNd birds heard a target mismatch during reinforcement whereas dLO birds did not. dC birds were not pitch reinforced, their analysis windows (letters as in Fig. 1)

matched those of manipulated birds in terms of time-since-deafening. (**B, C**) Pitch change between the last 2 h of reinforcement (R) and the pitch of successive days aligned to the first 2 h of song after withdrawal of reinforcement in std for WNd (red, B) and dLO (blue, C) birds. Early and late windows are marked with markers (dots, crosses) in B and letters (E, L) in C. Curves are plotted such that pitch changes towards the target are pointing down (see Methods). Example birds shown in D, E are marked with arrows.(**D, E**) Syllable pitches (dots, red=WN reinforced, blue=LO-reinforced syllables) of example WNd (D) and dLO (E) bird. (**F**) WNd (red) perform both early and late pitch changes in the direction of the baseline target (by about one standard deviation, * $p < 0.05$, *** $p < 0.001$, one-tailed t-test), similar to WNC (gray) and unlike dLO (blue) birds without mismatch experience. Dotted lines mark y-range displayed in (**G**) Bootstrapped pitch differences between reinforced WNd (red) and dLO (blue) and $10'000$ times randomly matched dC birds, shown for early (solid line) and late (dashed line) analysis windows. The stars indicate the bootstrapped probability of a zero average pitch difference between reinforced and dC birds (n.s. not significant, ** $p < 0.01$, *** $p < 0.001$).

In contrast, dLO birds showed no signs of reverting pitch, neither in the first 2 h since release of reinforcement ($d' = -0.13$, $p = 0.36$, $t$stat $= -0.37$, $df = 7$, $N = 8$ birds, one-sided t-test), nor after 4 days of practice ($d' = -0.08$, $p = 0.43$, tstat $= -0.18$, $df = 7$, $N = 8$ birds, one-tailed t-test, Fig. 4C, E, F).

The singing rate does not explain why deaf birds with mismatch experience partially revert their song towards baseline, unlike deaf birds without mismatch experience. WNd birds sang less during the first 2 h after reinforcement (early) than both control birds ($p = 2.3 \cdot 10^{-6}$, tstat $= 6.02$, $df = 26$, $N = 10$ WNd and $N = 18$ WNC birds, two-sided t-test) and dLO birds ($p = 0.008$, tstat $= 3.06$, $df = 16$, $N = 8$ dLO birds, two-sided t-test), unlike dLO birds that sang similar amounts as WNC birds ($p = 0.11$, tstat $= 1.67$, $df = 24$, two-sided t-test). If the number of songs were to determine the rate of recovery, we would have seen the opposite effect (dLO birds should recover similar amounts as WNC birds and significantly more than WNd birds). In conclusion, singing rate does not explain the difference between WNd and dLO birds.

To discount for the effect of time elapsed since deafening and quantify the change in pitch specifically due to reinforcement, we bootstrapped the difference in d' between dLO/WNd birds and a new group of dC birds that were deafened but experienced no prior reinforcement (see methods). To discount for possible influences of circadian pitch trends, we assessed early and late pitch changes in reinforced birds and in dC birds in 2 h time windows separated by multiples of 24 h (and again flipped pitch changes in birds that were reinforced to decrease pitch, see Methods). In agreement with the findings above, we found that significant reversion towards baseline was only seen in WNd birds and very consistently so (Fig. 4G, Table S1), showing that prior experience of a target mismatch is necessary for pitch reversion independent of auditory feedback.

We further validated our finding using a linear mixed effect model on the combined NRP data of all groups (see Methods), which confirmed our previous findings: We did not find a significant effect of the time without practice between R and E windows on the NRP in the E window (fixed effect -0.04, $p = 0.2$), confirming that birds do not recover without practice. Neither deafening nor muting had a significant effect by itself but the interaction

between deafening and time (late) was associated with an NRP increase of 0.67 (fixed effect, p=2*10^{-6}), demonstrating that deaf birds are significantly further away from baseline ($NRP = 0$) than hearing birds in late windows, thereby confirming that birds require auditory feedback to recover a distant pitch target. Importantly, we found that mismatch experience was associated with a significant fixed effect of -0.37 on the NRP (fixed effect towards the target, p=0.006), supporting our finding that limited vocal plasticity is possible even in the absence of auditory feedback.

Our results thus argue for a model of song maintenance in which birds extract from target mismatch experience a plan of reducing the mismatch. Without practice and auditory experience, birds cannot reach a distant motor target (Fig. 5A). With practice and without auditory experience, they can make small changes towards a target, which we refer to as the planning range. Auditory experience allows them to consolidate the small changes such that step-by-step they can reach even a distant target (Fig. 5B).

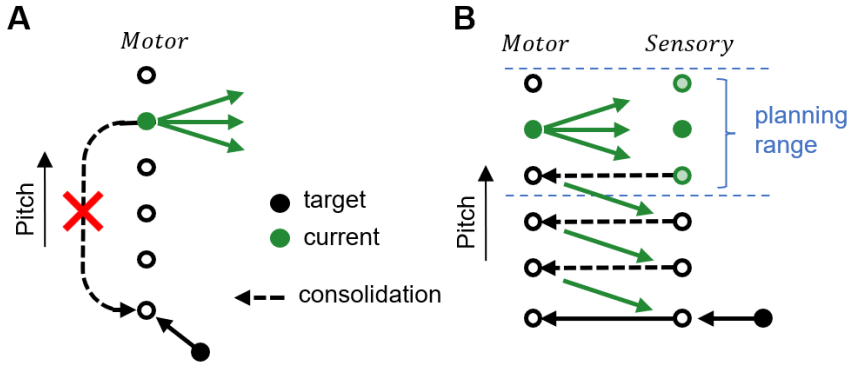

Fig. 5. Schematic illustrating the goal-directed planning of vocal changes. (**A**) Without practice, birds cannot recover a distant motor target (black filled circle) far away from the current motor output (green filled circle). (**B**) Without auditory experience, birds can make motor changes (green arrows) towards a target within a small range, we refer to this range as the (overt) planning range (blue). To recover a distant target (black filled circle) beyond the planning range, birds need auditory experience (green circles under Sensory), presumably to consolidate (dashed arrows) the overt motor changes.

Discussion

Our work shows that recent auditory experience can drive motor plasticity even while an individual is deprived of such experience, i.e. zebra finches are capable of overt vocal planning. But to reach a distant vocal target beyond the pitch range they have recently produced necessitates auditory feedback, which sets a limit to zebra finches' overt planning ability.

Our insights were gained in deaf birds and we cannot rule out that deaf birds could gain access to pitch information via somatosensory-proprioceptive sensory modalities. However, such information, even if available, cannot explain the difference between the "mismatch experience" (WNd) and the "no mismatch experience" (dLO) groups, which strengthens our claim that the pitch reversion we observe is a planned change and not merely a rigid motor response (as in simple use-dependent forgetting (5,6)). Also, it is unlikely that dLO birds' inability to recover baseline pitch is somehow due to our use of a

reinforcer of a non-auditory (visual) modality, since somatosensory stimuli do not prevent reliable target pitch recovery in hearing birds (42). Thus, the overt planning ability is an active experience-dependent process.

In our two-stage model, recovery of a developmentally learned vocal target is controlled by two nested processes, a highly flexible process with limited scope ($d' \simeq 1$, Fig. 4, 5), and a dependent process enabled by experience of the former. Such motor learning based on separate processes for acquisition and retention is usually referred to as motor consolidation (4,34,35). Accordingly, the flexible process of acquisition or planning as we find is independent of immediate sensory experience, but the dependent process (consolidation of the flexible process) requires experience. Perhaps then, it is the sensory experience itself that is consolidated, and therefore, consolidation of sensory experience may be a prerequisite for extensive planning.

We cannot distinguish the overt planning we find from a more complex use-and-experience dependent forgetting, since we only probed for recovery of pitch and did not attempt to push birds into planning pitch shifts further away from baseline. Evidence for more flexible planning is provided by the pitch matching skills of nightingales (3). Interestingly, although nightingales can reach without practice even distant pitch targets, the targets in (3) were also located within the extent of nightingale's recent song practice, so also satisfied $d' \simeq 1$. Perhaps then, our two-stage model of song plasticity of planning and consolidation in Fig. 5 applies more broadly in songbirds and not just in zebra finches.

Consolidation in motor learning generally emerges from anatomically separated substrates for learning and retention (5). Such separation also applies to songbirds. Both reinforcement learning of pitch and recovery of the original pitch baseline depend on the anterior forebrain pathway and its output, the lateral magnocellular nucleus of the anterior nidopallium (LMAN)(16). LMAN generates a pitch bias that lets birds escape negative pitch reinforcers and recover baseline pitch when reinforcement is withdrawn (14), thus is likely involved in planning. This pitch bias is consolidated outside of LMAN (16,36) in a nonlinear process that is triggered when the bias exceeds a certain magnitude (37). This threshold magnitude is roughly identical to the planning limit we find ($d' \simeq 1$), suggesting that birds' planning limit arises from the consolidation of LMAN-mediated motor plasticity. Although it remains to be seen whether LMAN is capable of executing motor plans without sensory feedback, our work provides a new perspective on the neural basis of birdsong learning and consolidation in and around LMAN.

The formation of a planned motor change may not require LMAN itself, because pharmacological suppression of LMAN sets the bias to zero, but upon removal of output suppression, the pitch of the song syllable that was targeted by reinforcement jumps by about 1% away from the reinforced pitch zone (38), which corresponds to about $d' = 1$, about the planning limit we find. Originally, this jump was interpreted as evidence of functional connectivity or an efference copy between the anterior forebrain pathway of which LMAN is part of and some other unspecified variability-generating motor area. However, in our view, a simpler explanation requiring neither functional connectivity nor efference copy is that LMAN is involved in putting a plan into action, which in that case is to produce syllable variants that are unaffected by WN.

Zebra finches' ability to plan directed song changes could hinge on song memories that feed into LMAN and that could drive neurons there to produce diverse perceptual song variants. LMAN neurons are selective for the bird's own song but not the target song

(21,22), which makes them well suited for executing song plans within the range of recent experience (i.e., if the song is outside recent experience, it elicits no LMAN response and so does not gain access to planning circuits). Furthermore, LMAN neurons show mirrored activity, i.e., similar activity when a zebra finch produces a vocal gesture and when it hears the same gesture played through a loudspeaker (39,40). This mirrored activity has been argued to be involved in translating an auditory target into the corresponding motor command, also known as an inverse model (41). Mirroring in LMAN was observed across the song variability generated over a period of several hours, which is about the same as the experience-dependent pitch planning limit we find. Zebra finches could thus transform a desired pitch change into the corresponding motor plan via LMAN's aligned sensory and motor representations of recent vocal output.

In a broader context of motor recovery, birds' failure to recover baseline pitch without guiding sensory feedback agrees with reports that binary reinforcement (as we used) slows down or prevents forgetting of the adapted behavior (6). However, whereas forgetting is fast when sensory errors affect arm movements (6), the contrary applies to birdsong, where pitch learning from artificial sensory errors is slower and less forgotten (42) than is pitch learning from binary reinforcement (13,15). Hence, the commonality of short-term visuo-motor adaptation and of birdsong maintenance is that slow learning leads to slow forgetting, regardless of whether it is due to sensory errors or reinforcement. Such conclusion also agrees with observations that zebra finch song does not recover to pre-manipulated forms, both after restoring auditory feedback after long-term (>5 months) deprivation (43) and after restoring normal syrinx function after long-term (16 weeks) manipulation with beads (44), suggesting that song can spontaneously recover only within some limited time since it was manipulated.

Our observations in zebra finches could be relevant to other species including humans. The planning abilities we find bear resemblance to human motor imagery for movement learning, which is most effective when subjects already show some competence for the movements to be learned (45), suggesting a recall-dependent process. Naively, human vocal flexibility seems superior to that of zebra finches, since we can flexibly change sound features such as loudness, pitch, and duration to convey emotional state or to comply with the tonal and rhythmical requirements of a musical piece (46,47), whereas zebra finches produce more subtle modulations of their songs e.g. when directing them to a female (48). Nevertheless, a limit of human vocal flexibility is revealed by non-native accents in foreign languages, which are nearly impossible to get rid of in adulthood. Thus, a seeming analogous task to feedback-dependent learning of zebra finch song, in humans, is to modify developmentally learned speech patterns.

Our findings help elucidate the meaning of song signals in songbirds and the evolutionary pressures of singing. Because zebra finches seem incapable of large jumps in performance without practice, their current song variants are indicative of the recent song history, implying that song is an honest signal that zebra finches cannot adapt at will to deceive a receiver of this signal. Hence, if high pitch has either an attractive or repelling effect on another bird, a singer must commit to being attractive or repulsive for some time. In extension, we speculate that limited vocal flexibility increases the level of commitment to a group and thereby strengthens social cohesion.

Materials and Methods

All experimental procedures were in accordance with the Veterinary Office of the Canton of Zurich (licenses 123/2010 and 207/2013) or by the French Ministry of Research and the ethical committee Paris-Sud and Centre (CEEA N°59, project 2017-12).

***Subjects*:**

We used in total 76 birds. All birds were 90-300 days old (except one 853-day old control bird) and were raised in the animal facility of the University of Zurich or in Saclay. During recording, birds were housed in single cages in custom-made sound-proof recording chambers equipped with a wall microphone (Audio-Technica Pro4 and 2), a loudspeaker. The day/night cycle was set to 14/10 h except for one muted bird that was in constant light due to a technical problem.

Song Recordings:

Vocalizations were saved using custom song-recording software (Labview, National Instruments Inc.). Sounds were recorded with a wall microphone and digitized at 32 kHz. In all birds, we recorded baseline vocal activity for at least 3 days before doing any manipulation (deafening or pitch reinforcement).

Pitch Reinforcement:

We calculated pitch (fundamental frequency) as described in (15). To provide pitch reinforcement in real time, we used a two-layer neural network trained to detect a manually clustered syllable containing a harmonic stack (31). We evaluated the fundamental frequency of that syllable in a 16-24-ms time window following detection. For pitch reinforcement, we either broadcast a 50-60-ms long white noise (WN) stimulus through a loudspeaker or briefly switched off the light in the isolation chamber for 100-500 ms (LO) when pitch was below or above a manually set threshold. The WN/LO stimulus onset occurred 7 ms after the pitch calculation offset. We performed cumulative pitch shifts across several days by adjusting the pitch threshold for WN/LO delivery each day, usually setting it close to the median value of the previous day. Sometimes the threshold was set more than once during a day, in this case we set it close to the median of the pitch values measured so far during that day. All birds were shifted by at least 1 standard deviation ($d' > 1$, see Section Pitch Analysis).

Reported pitch values were collected as above, except in muted birds that directly after unmuting produced syllables of lower amplitude and with distorted spectral features (e.g. Fig. 1C, E, F), which resulted in frequent mis-detections by the neural network. In muted birds, we therefore performed semi-automatic (manually corrected) syllable detection and we computed pitch at a fixed time lag after syllable onset. Despite deafening leading to degradation of birds' song (49), syllable detection and pitch calculation were still possible in all deaf birds (birds were recorded during 13-50 days after deafening surgery, age range 90-300 dph, n=44 birds). Since pitch shifting was balanced in all deaf bird groups (the same number of birds were up- and down-shifted), systematic changes in pitch post deafening (49) will average out and so would not affect our findings.

Duration Reinforcement:

Duration reinforcement was performed similarly as pitch reinforcement but instead of measuring the pitch of a targeted syllable, we measured the duration of a targeted song element (either a syllable, a syllable plus the subsequent gap, or just a gap). Onsets and offsets of the targeted element were determined by thresholding of the root-mean square (RMS) sound amplitude.

Bird groups:

WN Control (WNC)*:* 18 birds in the control group underwent WN pitch reinforcement (10/18 up-shifted, 8/18 down-shifted). Thereafter, the WN stimulus was withdrawn, and no further experimental manipulation took place.

WN muted (WNm)*:* in 8 birds, we first reinforced pitch using white noise (WN) auditory stimuli and then we reversibly muted the birds by performing an airsac cannulation. Normally, when WN stimuli are contingent on low-pitch renditions, birds tend to shift the pitch up, and in 5/6 birds this was indeed the case. However, one bird shifted the pitch down, in an apparent appetitive response to WN, this bird responded appetitively also when the WN contingency was changed, resulting in a net upward shift at the end of the WN period, see also (31). In 2 birds, we targeted high-pitch variants and these birds shifted the pitch down, as expected. Thus, in total, in 6/8 birds (including the bird with the apparent appetitive response), we drove the pitch up and in 2/8 birds, we drove the pitch down.

Two birds underwent the muting surgery directly after withdrawal of WN stimuli. We observed that the two birds had recovered a mere 10% and -6% of their total WN-driven pitch change (Fig. 1H). We hypothesized that unreinforced singing would initiate the song recovery process in WNm birds that we assumed birds might be able to accomplish while mute. Therefore, we allowed the subsequent 6/8 WNm birds (4 up-shifted and 2 down-shifted) to sing a few hundreds of target syllables without reinforcement prior to muting them. During on average 4h51mins (range 10 mins to 14h), these latter birds produced on average 649 song motifs (56, 100, 400, 458, 480, and 2400 motifs) without WN; the example bird shown in Fig. 1C produced 56 song motifs within 11 minutes during the 30 minutes it was allowed to sing without aversive reinforcement.

WN deaf (WNd)*:* 10 birds were first pitch reinforced (5/10 were up-shifted and 5/10 down-shifted) with WN and then they were deafened immediately after by bilateral cochlea removal. WNd birds started to sing on average 3±1 days after deafening (range 2 to 5 days) and were recorded for at least 15 days after the deafening surgery.

Deaf LO (dLO)*:* 8/10 birds from (32) were recorded after the reinforcement period and we analyzed the associated data. These birds were first deafened by bilateral cochlea removal, then they underwent pitch reinforcement with light-off (LO) stimuli that acts as an appetitive stimulus in deaf birds. The lamp in the recording chamber was switched off for 100-500 ms when the pitch was either above or below a manually set threshold (daily threshold adjustment followed the same procedure as for WNm birds). 3/8 birds received LO for low-pitched syllables and 5/8 birds for high-pitched syllables. One of the birds that received LO for high-pitched syllables changed its pitch away from LO instead of towards it, thus we ended up with a balanced data set with 4/8 birds shifting pitch up and 4/8 birds shifting down. dLO birds were recorded for at least 5 days after the deafening surgery. Details of light-induced pitch shifting are described in (32).

Deaf control (dC)**:** we analyzed 26 syllables from 20 birds taken (12 from (32) and 8 additional ones) that were deafened and then recorded without any further manipulation. We used these birds to discount for pitch changes in WNd and dLO birds due to absence of auditory feedback, see bootstrapping.

WN duration (WNdur): 12 birds underwent duration reinforcement using WN, in 9 birds the targeted sound feature was syllable duration, in 2 birds the targeted feature was syllable-plus-gap duration, and in one bird the targeted feature was gap duration. In 4 birds, the duration was squeezed and in 8 birds the duration was stretched. As in WNC birds, we did no further experimental manipulation after withdrawal of the WN stimulus.

One bird changed its duration towards WN showed an apparent appetitive response to WN as for the one muted bird.

Muting

We muted birds by inserting a by-pass cannula into the abdominal air sac (50) as follows.
Preparation of by-pass cannula: After incubation in 70% ethanol, we clogged a 7 mm long polyimide tube (diameter 1.2 mm) with sterile paper tissue. We created a suture loop around the cannula and fixed the thread to the cannula with a knot and a drop of tissue glue.
Cannula implantation: We anaesthetized the birds with Isoflurane (1.5-2%) and gave a single injection of Carprofen (4 mg/kg). Subsequently, we applied local analgesic to the skin (2% lidocaine) and removed the feathers covering the right abdomen. We applied Betadine solution on the exposed skin and made a small incision using sterilized scissors. We exposed the right abdominal air sac by shifting aside the fat tissue and punctured it to create an opening. Immediately, we closed the opening by inserting the cannula and by sealing the contact region with tissue glue. With the free end of the glued thread, we made one suture to the lowest rib. We closed the wound in the skin around the cannula with tissue glue and sutures using a new thread. Finally, we applied betadine solution on the wound and lidocaine gel around the injured site. Before releasing the bird to its cage, we removed the clog of the cannula with forceps and verified the air flow through the cannula.
We returned the birds to their home cage and monitored them for signs of suffering. We administered pain killers (Meloxicam 2 mg/kg or Carprofen 2-4 mg/kg) for 2 days after the surgery.
On the following days, we monitored the birds continuously for singing activity. If song was detected, the cannula was inspected for clogging and cleaned. 5 birds unmuted spontaneously, they produced at most 300 songs before the bypass cannula was inspected and the clog was removed to re-mute the bird. To unclog the bypass cannulas, we used sharp forceps and sterile tissue dipped in saline. 6 of 8 birds produced quiet call-like vocalizations even on muted days on which no singing was detected.

Deafening

We bilaterally removed cochleas as described in (32).

Pitch Analysis

In individual birds, we studied the dynamics of pitch recovery during the test period. In WNm birds, the test period started with unmuting, and in all other reinforced birds it started with the end of reinforcement. We analyzed songs in early (E) time windows defined as the first 2 h window during the test period in which the bird produced at least 20 song motifs. We also assessed pitch recovery in late (L) windows defined exactly 4 days after the E window. To make the measurements robust to circadian fluctuations of pitch, we compared the pitch values in early and late windows to pitch values produced in time-aligned windows during the last day of reinforcement (R) and during the last day of baseline (B).
We used this time-of-day matched analysis to produce Fig. 1H, I, Fig. 2. C, D and Fig. 3. C, D. Exceptions where time alignment was not possible are listed in the following:

- One WNm bird started singing late on the last day of reinforcement (preventing us from time-aligning the R window with the E window), and therefore in this bird we defined R after the end of WN but before muting (in this bird there is more than one day of song after WN and before muting).

- In two birds (1 WNC and 1 dLO bird), we defined the L window one day earlier (on the 4th day, after 3 days of practice), because there was no data for these birds on the 5th day after reinforcement (our findings did not qualitatively change when we defined the L window on the 6th day instead of the 4th).
- One WNm bird was housed together with a female during WN reinforcement; this bird did not sing during the time-match 2-h period on the 2nd, 3rd, and 4th day after reinforcement; therefore on those days we computed the mean pitch from all values produced on that day in Fig. 1H.

In early (E) and late (L) analysis windows, we computed the normalized residual pitch (NRP), which is the remaining fraction of pitch shift since release from WN, defined as $NRP(X) = (P_X - P_B)/(P_R - P_B)$, where P_X is either the mean pitch in the early $(X = E)$ or late $(X = L)$ window (Fig 1H, I, 2C, D, 3C, D). P_R and P_B are the mean pitches in the R and B windows, respectively. An NRP of 33% indicates that two-thirds of the reinforced pitch shift have been recovered and an NRP of 0% indicates full recovery of baseline pitch. Note that the NRP measure discounts for differences in the amount of initial pitch shift the birds displayed at the beginning of the test period.
We performed statistical testing of NRP to discount for this diversity in initial pitch. To test the hypothesis that WNm birds recovered their baseline pitch without practice or that WNd or dLO birds recovered baseline pitch without auditory feedback, we performed a two-tailed t-test for $NRP = 0$.
Our results were qualitatively unchanged when we changed the timing of the L window, as long as there were at least 3 days between E and L windows (because WNC birds need at least 3 days to recover their baseline pitch in the L window, $p < 0.05$). Thus, giving deaf birds more time did not allow them to recover their baseline pitch. Furthermore, we also tested larger windows of 4 and 24 h duration instead of 2 h and found qualitatively similar results. We further verified that our results did not critically depend on the time-alignment by repeating the NRP tests using the last 2 h of reinforcement as the R windows. Indeed, we found that all results in Fig. 1-3 were unchanged.
We computed the pitch change after reinforcement (Fig. 4) as the difference in mean pitches between early (E) or late (L) and the last 2 h of WN/LO reinforcement R in units of sensitivity $d' = (P_X - P_R)/S_R$, where S_R is the standard deviation of pitch values in the R window. To test the hypothesis that WNd and dLO birds are able to make targeted pitch changes towards baseline, we performed a one-tailed t-test of the hypothesis H0: $d' < 0$. We used sensitivity d' relative to the last 2 h of WN/LO instead of NRP because we want to detect a pitch change, which is the realm of detection theory, i.e. d'. Furthermore, by measuring local changes in pitch relative to the last 2 h of WN/LO reinforcement, our measurements are only minimally affected by the amount of reinforcement learning that might have occurred during this 2 h time window — choosing an earlier or longer window would have blended reinforced pitch changes into our estimates. Last but not least, changes in the way in which we normalized d' values — dividing by S_B, $\sqrt{(S_B^2 + S_R^2)/2}$, or $\sqrt{(S_X^2 + S_R^2)/2}$ — or using the NRP relative to the last 2 h of WN/LO did not qualitatively change the results shown in Fig. 4D.

Bootstrapping

To test whether deaf birds indeed make small pitch changes towards a target if and only if they experienced target-mismatch during reinforcement, we bootstrapped the difference in pitch changes between reinforced (WNd and dLO) and deaf control birds (dC). All dC birds were recorded for at least 5 days after they started singing while deaf.

In dC birds, we defined the R, E, and L windows such that they matched those of WNd and dLO birds in terms of days since deafening. Additionally, in dLO birds we chose the windows such that they matched in terms of time-of-day (because LO always ended overnight). Thus, the R windows in dC birds either corresponded to the last 2 h before deafening (as control for WNd birds) or to the last 2 h of the day before E (as control for dLO birds).

For WNd birds, we obtained in total 26 control syllables from 20 dC birds. For dLO birds, we obtain 17 control syllables from 13 dC birds (some dC birds did not provide any useable data because they stopped singing or were not recorded for long enough).

For the bootstrapping procedure, we randomly paired control syllables (N=26 for WNd and N=17 for dLO) one-by-one with matchable syllables from reinforced birds (with replacement), computed the mean pitches P_R , P_E, P_L in corresponding windows, calculated the standard deviation S_R, calculated the average pitch changes $\mathrm{d}'_E = (P_E - P_R)/S_R$ and $\mathrm{d}'_L = (P_L - P_R)/S_R$ for both manipulated and control birds, and multiplied these by -1 if the reinforced bird was down-shifted (as we did for d' above). We then took the differences in average pitch changes between manipulated (WNd and dLO) and dC birds, e.g. $\mathrm{d}'_{E,\mathrm{WNd}} - \mathrm{d}'_{E,\mathrm{dC}}$. We repeated this procedure 10'000 times and plotted the distribution of average pitch change differences between WNd and dC (red) and between dLO and dC (blue) in Fig. 4E and perform bootstrap statistics.

Our results were qualitatively unchanged (only WNd significantly reverted pitch towards baseline) when we aligned the R windows by the time-of-day of the corresponding E windows (two dC birds started singing later on the day of the E window than they stopped singing on the days before; in these two birds we used the R windows instead), see Table S1. Although the d' values in both groups increased (and in dLO birds, the average d' in the L windows was positive, $p < 0.05$, two-tailed t-test), we found a significant pitch difference between WNd or dLO birds in L windows, which upholds our findings that mismatch experience is necessary for pitch reversion. The reason for the increases in d' likely is that birds further shifted their pitch away from baseline on the last day of reinforcement (after the time-aligned R window). Also, results were robust when we analyzed pitch changes after release from reinforcement in units of NRP: without practice, WNd birds made small and significant pitch changes towards baseline, and dLO birds stayed at $NRP \geq 1$.

Linear Mixed Effect Model

We simulated a linear mixed effect model on the combined NRP data from all groups with fixed effects corresponding to time (general offset a, late b), treatment (deafened c, muted d), mismatch-experience e, and a fixed effect f that is linear in the time between the time $t(R)$ of the R window and the time $t(E)$ of the E window. We used this latter term to test whether birds can recover without practice. We further included a fixed effect in terms of the interaction g between deafening and late, to test whether birds recover without auditory feedback (but with practice from E to L):

$$NRP_{i,t} = a + b\delta_{t\in L} + c\delta_{i\in\mathrm{WNd\ or}\ i\in\mathrm{dLO}} + d\delta_{i\in WNm} + e\delta_{i\notin\mathrm{dLO}} + f\big(t(E) - t(R)\big)\delta_{t\in E} + g\delta_{i\in\mathrm{WNd\ or}\ i\in\mathrm{dLO}}\delta_{i\in\mathrm{L}} + \varepsilon_{i|\mathrm{group}}$$

The Kronecker δ of a specific group equals 1 if bird i belongs to that group (e.g. $i \in WN$) resp. if the time window t is either E or L (e.g. $t \in L$), and it equals 0 otherwise. The

terms $\varepsilon_{i|\text{group}}$ is a random effect associated with a particular bird i and group. Note that the fixed effect of mismatch experience e is zero for dLO birds and it is one for all other birds.

References

1. Brecht KF, Hage SR, Gavrilov N, Nieder A. Volitional control of vocalizations in corvid songbirds. PLoS Biol. 2019 Aug 27;17(8):e3000375.
2. Veit L, Tian LY, Monroy Hernandez CJ, Brainard MS. Songbirds can learn flexible contextual control over syllable sequencing. Elife. 2021 Jun 1;10.
3. Costalunga G, Carpena CS, Seltmann S, Benichov JI, Vallentin D. Wild nightingales flexibly match whistle pitch in real time. Curr Biol. 2023 Aug 7;33(15):3169–3178.e3.
4. Brashers-Krug T, Shadmehr R, Bizzi E. Consolidation in human motor memory. Nature. 1996 Jul 18;382(6588):252–255.
5. Galea JM, Vazquez A, Pasricha N, de Xivry J-JO, Celnik P. Dissociating the roles of the cerebellum and motor cortex during adaptive learning: the motor cortex retains what the cerebellum learns. Cereb Cortex. 2011 Aug;21(8):1761–1770.
6. Shmuelof L, Huang VS, Haith AM, Delnicki RJ, Mazzoni P, Krakauer JW. Overcoming motor “forgetting” through reinforcement of learned actions. J Neurosci. 2012 Oct 17;32(42):14617–14621.
7. Krakauer JW, Ghez C, Ghilardi MF. Adaptation to visuomotor transformations: consolidation, interference, and forgetting. J Neurosci. 2005 Jan 12;25(2):473–478.
8. Sheahan HR, Franklin DW, Wolpert DM. Motor planning, not execution, separates motor memories. Neuron. 2016 Nov 23;92(4):773–779.
9. Konishi M. Birdsong: from behavior to neuron. Annu Rev Neurosci. 1985;8:125–170.
10. Immelmann K. Song development in the zebra finch and other estrildid finches. Bird vocalizations. 1969;
11. Park S-W, Dijkstra TMH, Sternad D. Learning to never forget-time scales and specificity of long-term memory of a motor skill. Front Comput Neurosci. 2013 Sep 2;7:111.
12. Smith MA, Ghazizadeh A, Shadmehr R. Interacting adaptive processes with different timescales underlie short-term motor learning. PLoS Biol. 2006 Jun;4(6):e179.
13. Tumer EC, Brainard MS. Performance variability enables adaptive plasticity of “crystallized” adult birdsong. Nature. 2007 Dec 20;450(7173):1240–1244.
14. Andalman AS, Fee MS. A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. Proc Natl Acad Sci USA. 2009 Jul 28;106(30):12518–12523.
15. Canopoli A, Herbst JA, Hahnloser RHR. A higher sensory brain region is involved in reversing reinforcement-induced vocal changes in a songbird. J Neurosci. 2014 May 14;34(20):7018–7026.
16. Warren TL, Tumer EC, Charlesworth JD, Brainard MS. Mechanisms and time course of vocal learning and consolidation in the adult songbird. J Neurophysiol. 2011 Oct;106(4):1806–1821.

17. Hoffmann LA, Saravanan V, Wood AN, He L, Sober SJ. Dopaminergic contributions to vocal learning. J Neurosci. 2016 Feb 17;36(7):2176–2189.

18. Aronov D, Andalman AS, Fee MS. A specialized forebrain circuit for vocal babbling in the juvenile songbird. Science. 2008 May 2;320(5876):630–634.

19. Nottebohm F, Stokes TM, Leonard CM. Central control of song in the canary, Serinus canarius. J Comp Neurol. 1976 Feb 15;165(4):457–486.

20. Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R. Neural correlates of categorical perception in learned vocal communication. Nat Neurosci. 2009 Feb;12(2):221–228.

21. Yazaki-Sugiyama Y, Mooney R. Sequential learning from multiple tutors and serial retuning of auditory neurons in a brain area important to birdsong learning. J Neurophysiol. 2004 Nov;92(5):2771–2788.

22. Kojima S, Doupe AJ. Song selectivity in the pallial-basal ganglia song circuit of zebra finches raised without tutor song exposure. J Neurophysiol. 2007 Oct;98(4):2099–2109.

23. Yanagihara S, Yazaki-Sugiyama Y. Auditory experience-dependent cortical circuit shaping for memory formation in bird song learning. Nat Commun. 2016 Jun 21;7:11946.

24. Yu K, Wood WE, Theunissen FE. High-capacity auditory memory for vocal communication in a social songbird. Sci Adv. 2020 Nov 13;6(46).

25. Riters LV, Spool JA, Merullo DP, Hahn AH. Song practice as a rewarding form of play in songbirds. Behav Processes. 2019 Jun;163:91–98.

26. Kollmorgen S, Hahnloser RHR, Mante V. Nearest neighbours reveal fast and slow components of motor learning. Nature. 2020 Jan 8;577(7791):526–530.

27. Kojima S, Doupe AJ. Social performance reveals unexpected vocal competency in young songbirds. Proc Natl Acad Sci USA. 2011 Jan 25;108(4):1687–1692.

28. Elemans CPH, Rasmussen JH, Herbst CT, Düring DN, Zollinger SA, Brumm H, et al. Universal mechanisms of sound production and control in birds and mammals. Nat Commun. 2015 Nov 27;6:8978.

29. Galea JM, Mallia E, Rothwell J, Diedrichsen J. The dissociable effects of punishment and reward on motor learning. Nat Neurosci. 2015 Apr;18(4):597–602.

30. Suthers RA, Goller F, Wild JM. Somatosensory feedback modulates the respiratory motor program of crystallized birdsong. Proc Natl Acad Sci USA. 2002 Apr 16;99(8):5680–5685.

31. Yamahachi H, Zai AT, Tachibana RO, Stepien AE, Rodrigues DI, Cavé-Lopez S, et al. Undirected singing rate as a non-invasive tool for welfare monitoring in isolated male zebra finches. PLoS One. 2020 Aug 10;15(8):e0236333.

32. Zai AT, Cavé-Lopez S, Rolland M, Giret N, Hahnloser RHR. Sensory substitution reveals a manipulation bias. Nat Commun. 2020 Nov 23;11(1):5940.

33. Zai AT, Cavé-Lopez S, Rolland M, Giret N, Hahnloser RHR. Sensory substitution reveals a manipulation bias (Data set). ETH Research Collection; 2020.

34. Fenn KM, Nusbaum HC, Margoliash D. Consolidation during sleep of perceptual learning of spoken language. Nature. 2003 Oct 9;425(6958):614–616.

35. Karni A, Sagi D. The time course of learning a visual skill. Nature. 1993 Sep 16;365(6443):250–252.

36. Tian LY, Brainard MS. Discrete Circuits Support Generalized versus Context-Specific Vocal Learning in the Songbird. Neuron. 2017 Dec 6;96(5):1168–1177.e5.

37. Tachibana RO, Lee D, Kai K, Kojima S. Performance-Dependent Consolidation of Learned Vocal Changes in Adult Songbirds. J Neurosci. 2022 Mar 9;42(10):1974–1986.

38. Charlesworth JD, Warren TL, Brainard MS. Covert skill learning in a cortical-basal ganglia circuit. Nature. 2012 May 20;486(7402):251–255.

39. Hanuschkin A, Ganguli S, Hahnloser R. A Hebbian learning rule gives rise to mirror neurons and links them to control theoretic inverse models. frontiersin.org.

40. Oztop E, Kawato M, Arbib MA. Mirror neurons: functions, mechanisms and models. Neurosci Lett. 2013 Apr 12;540:43–55.

41. Giret N, Kornfeld J, Ganguli S, Hahnloser RHR. Evidence for a causal inverse model in an avian cortico-basal ganglia circuit. Proc Natl Acad Sci USA. 2014 Apr 22;111(16):6063–6068.

42. Sober SJ, Brainard MS. Adult birdsong is actively maintained by error correction. Nat Neurosci. 2009 Jul;12(7):927–931.

43. Zevin JD, Seidenberg MS, Bottjer SW. Limits on reacquisition of song in adult zebra finches exposed to white noise. J Neurosci. 2004 Jun 30;24(26):5849–5862.

44. Hough GE, Volman SF. Short-term and long-term effects of vocal distortion on song maintenance in zebra finches. J Neurosci. 2002 Feb 1;22(3):1177–1186.

45. Mulder T, Zijlstra S, Zijlstra W, Hochstenbach J. The role of motor imagery in learning a totally novel movement. Exp Brain Res. 2004 Jan;154(2):211–217.

46. Dichter BK, Breshears JD, Leonard MK, Chang EF. The control of vocal pitch in human laryngeal motor cortex. Cell. 2018 Jun 28;174(1):21–31.e9.

47. Belyk M, Brown S. Pitch underlies activation of the vocal system during affective vocalization. Soc Cogn Affect Neurosci. 2016;11(7):1078–1088.

48. Stepanek L, Doupe AJ. Activity in a cortical-basal ganglia circuit for song is required for social context-dependent vocal variability. J Neurophysiol. 2010 Nov;104(5):2474–2486.

49. Lombardino AJ, Nottebohm F. Age at deafening affects the stability of learned song in adult male zebra finches. J Neurosci. 2000 Jul 1;20(13):5054–5064.

50. Nilson PC, Teramitsu I, White SA. Caudal thoracic air sac cannulation in zebra finches for isoflurane anesthesia. J Neurosci Methods. 2005 Apr 30;143(2):107–115.

Acknowledgments

We thank Manon Rolland and Sophie Cavé-Lopez for performing some of the deafening surgeries and their support with the experiments, and Heiko Hörster for providing excellent animal breeding and animal care services.

Funding:

Swiss National Science Foundation (Projects 31003A_182638 and 31003A_156976/1)
European Research Council (ERC) Advanced Grant (268911, VOTECOM)

Author contributions:

Conceptualization: ATZ, AES, RHRH
Investigation: ATZ, AES, NG
Data Curation: ATZ, AES
Formal analysis : ATZ, RHRH
Visualization: ATZ, AES, RHRH
Funding acquisition: NG, RHRH
Supervision: NG, RHRH
Writing – original draft: ATZ, AES, RHRH
Writing – review & editing: ATZ, AES, NG, RHRH

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Pitch and duration data that support the findings of this study together with the MATLAB scripts to reproduce the analysis and figures will be made available at the ETH Research Collection upon publication of the article. The raw data underlying the pitch measurement is not deposited due to its size but is available from the authors upon reasonable request.