

# Goal-directed vocal planning in a songbird

## Authors

Anja T. Zai<sup>1,2\*</sup>, Anna E. Stepien<sup>1,2</sup>, Nicolas Giret<sup>3</sup>, Richard H.R. Hahnloser<sup>1,2</sup>

## Affiliations

<sup>1</sup>Institute of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland.

<sup>2</sup>Neuroscience Center Zurich (ZNZ), University of Zurich and ETH Zurich, Zurich, Switzerland.

<sup>3</sup>Institut des Neurosciences Paris Saclay, UMR 9197 CNRS, Université Paris Saclay.

\* corresponding author: [zaia@ethz.ch](mailto: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

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

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

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 When the reinforcer is withdrawn, birds recover their  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

## 77 78 79 **Results**

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

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

94 singing. Physical absence of such oscillations essentially strips muted birds from all pitch  
95 experience.

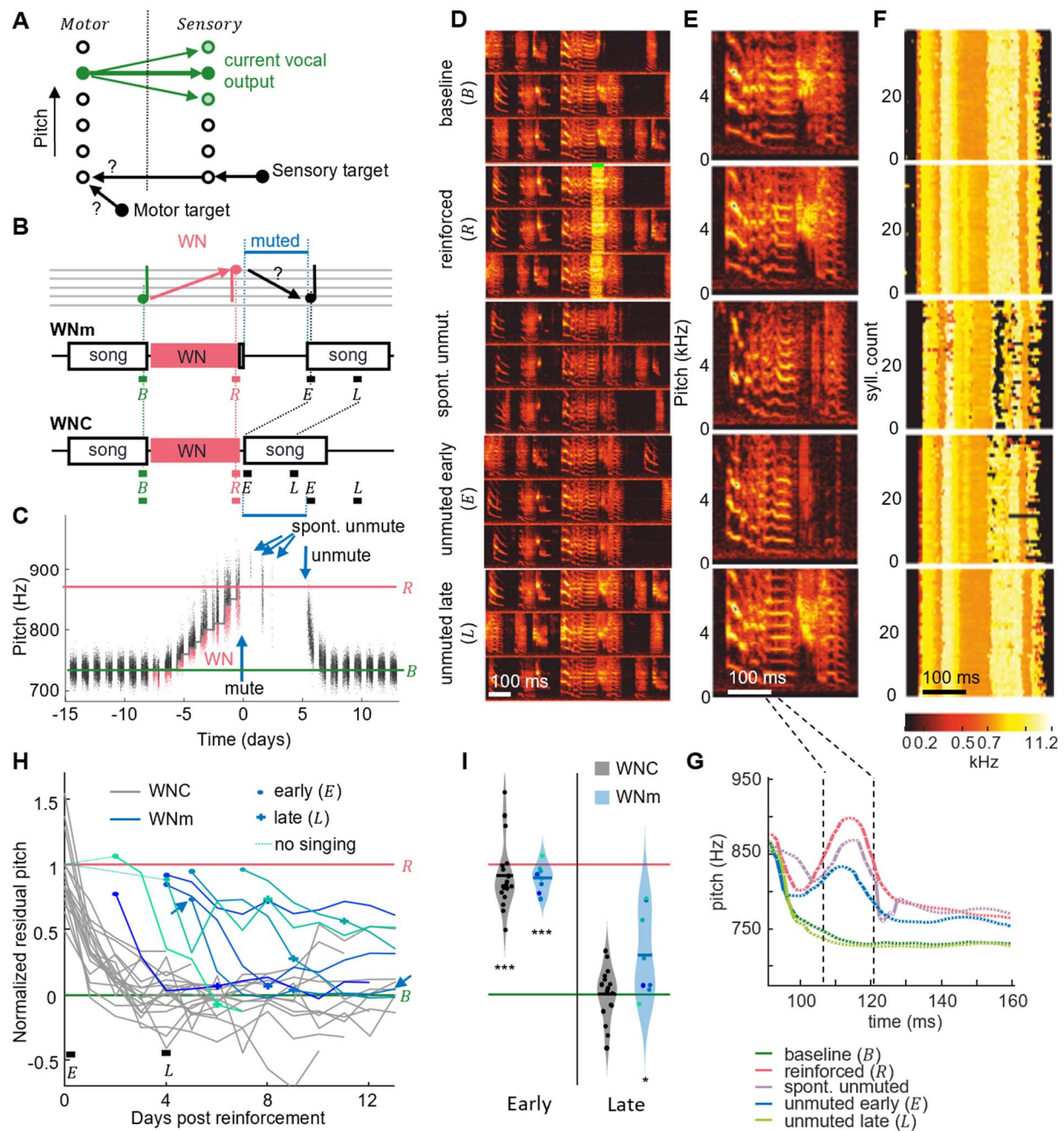
96  
97 We muted two birds right after the WN-driven pitch change. After keeping birds for four  
98 days in the muted state, we permanently unmuted them to record their undisturbed songs.  
99 We observed that the two birds had recovered a mere 10% and -6% of their total WN-  
100 driven pitch change (Fig. 1H). We hypothesized that unreinforced singing would initiate  
101 the song recovery process in WNm birds that we assumed birds might be able to  
102 accomplish while mute. Therefore, we allowed the subsequent 6/8 WNm birds to sing a  
103 few hundreds of target syllables without reinforcement prior to muting them.

104 In some cases, the bypass cannula got clogged during the muted period and birds were  
105 spontaneously unmuted, allowing them to produce a few songs before we reopened the  
106 cannula (Fig. 1C-G). These spontaneous unmuting events were not detrimental to our  
107 experimental procedures, as they allowed us to inspect birds' current song motor program  
108 (Fig. 1C).

109 After spending  $5.1 \pm 1.6$  days (range 3 – 8 d,  $N = 8$ ) in the muted state and upon  
110 unmuting, WNm birds displayed an average normalized residual pitch (*NRP*) of 89%,  
111 which was far from baseline ( $p = 6.2 \cdot 10^{-8}$ ,  $t_{\text{stat}} = -23.6$ ,  $N = 8$  birds, two-sided t-test  
112 of  $H_0: NRP = 0\%$ , songs analyzed in 2 h time window – early (*E*), see methods, Fig. 1),  
113 suggesting that in the muted state, birds are unable to recover their pre-reinforced songs.  
114 The average *NRP* in WNm birds was comparable to that of unmanipulated control (WNC)  
115 birds within the first 2 h after withdrawal of the reinforcer (average *NRP* = 91%,  $p =$   
116  $3.7 \cdot 10^{-11}$ ,  $t_{\text{stat}} = 14.8$ , two-sided t-test for *NRP* = 0%,  $N = 18$  WNC birds). Indeed,  
117 during 5 days without song practice, birds recovered no more pitch distance than birds  
118 normally do within the first 2 h of release from reinforcement ( $p = 0.82$ ,  $t_{\text{stat}} = -0.23$ ,  
119  $N = 8$  WNm and  $N = 18$  WNC birds, two-sided t-test). In WNm birds, there was no  
120 correlation between the *NRP* in the early window and the time since the muting surgery  
121 (correlation coefficient = 0.26,  $p = 0.53$ ), suggesting that the lack of pitch recovery  
122 while muted was not due to a lingering burden of the muting surgery. These findings did  
123 not sensitively depend on the size of the analysis window — we also tested windows of 4  
124 and of 24 h.

125 Subsequently, after 4 days of unmuted singing experience (roughly 9 days after  
126 withdrawal of WN), WNm birds displayed an average *NRP* of 30%, which was  
127 significantly different from the average *NRP* within the first 2 h after unmuting ( $p = 3 \cdot$   
128  $10^{-4}$ ,  $t_{\text{stat}} = 4.83$ ,  $N = 8$  birds, two-tailed t-test early (*E*) vs. late (*L*) time window) but  
129 still significantly different from zero ( $p = 0.04$ ,  $t_{\text{stat}} = 2.59$ ,  $N = 8$  birds, two-tailed t-  
130 test, late (*L*) time window). The amount of recovery was neither correlated with the  
131 number of renditions sung between early and late windows (correlation coefficient =  
132 0.03,  $p = 0.95$ ), nor with the duration the birds were muted (correlation coefficient =  
133  $-0.50$ ,  $p = 0.20$ ), nor with the time since they last sung the target song before  
134 reinforcement (correlation coefficient =  $-0.43$ ,  $p = 0.29$ ), suggesting the limiting factor  
135 for recovery was neither the amount of song practice nor the recovery time from the  
136 muting surgery (although for the latter there was a trend). Overall, these findings rule out  
137 covert planning in muted birds and suggest that motor practice is necessary for recovery of  
138 baseline song.

140



141

142

143

144

145

146

147

148

149

150

151

152

153

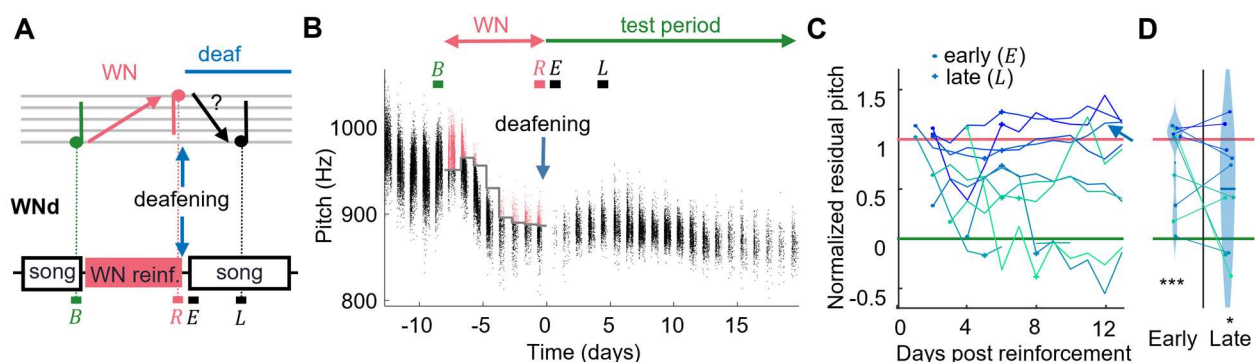
154

**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

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

170 Next, we tested whether motor experience but not sensory experience is necessary for  
 171 overt recovery, similar to arm reaching movements that can be restored without guiding  
 172 feedback (5,29). In a second group of birds, we provided slightly more singing experience  
 173 (Fig. 2). Instead of muting, WNd birds were deafened through bilateral cochlea  
 174 removal. This latter manipulation does not suppress the act of singing as does muting, but it  
 175 eliminates auditory feedback from singing. Deaf birds could gain access to some pitch  
 176 information via somatosensory stretch and vibration receptors and/or air pressure sensing  
 177 (30). Our aim was to test whether such putative pitch correlates are sufficient for recovery  
 178 of baseline pitch (Fig. 2A). However, in the deaf state, WNd birds did not recover baseline  
 179 pitch even after 4 days of song practice: on the 5<sup>th</sup> day (late, L) after deafening, their  
 180 average NRP was still 50%, which was different from zero ( $p = 0.03$ ,  $t_{stat} = 2.73$ , two-  
 181 tailed t-test of  $H_0: NRP = 0\%$ ,  $N = 10$ , Fig. 2D) and significantly larger than the average  
 182 NRP of WNC birds on the 5<sup>th</sup> day since withdrawal of reinforcement (difference in  
 183  $NRP = 49\%$ ,  $p = 0.003$ ,  $t_{stat} = 3.34$ ,  $df = 26$ ,  $N = 10$  WNd and  $N = 18$  WNC birds,  
 184 two-tailed t-test).

185



186

187

188

189

190

191

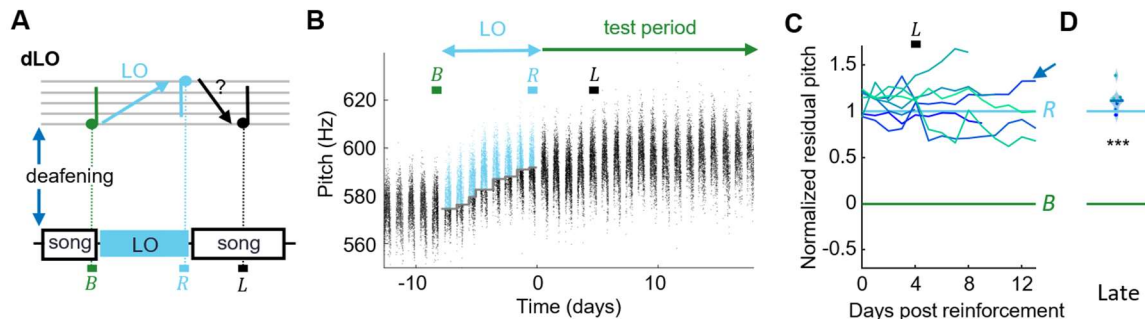
192

193

**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,

194 cross=late). (D) Violin plots of same data restricted to early and late analysis  
195 windows, lines connect individual birds ( $***p < 0.001$ ,  $*p < 0.05$ , two-tailed t-  
196 test of  $NRP = 0$ ).

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



204  
205 **Fig. 3. Deaf birds do not recover pitch target after light-induced mismatch.** (A) dLO  
206 birds were first deafened and then pitch-reinforced using a brief light-off (LO)  
207 stimulus. Analysis windows (letters) as in Fig. 1. (B) Syllable pitches (dots,  
208 blue=LO-reinforced syllables) of example dLO bird that shifted pitch up by  $d' =$   
209  $3.5$  within a week, but showed no signs of pitch recovery during the test period.  
210 (C) dLO birds ( $N = 8$ ) do not recover baseline pitch without auditory feedback.  
211 (D) Violin plots of same data restricted to the late analysis window ( $***p <$   
212  $0.001$ , two-tailed t-test of  $NRP = 0$ ).

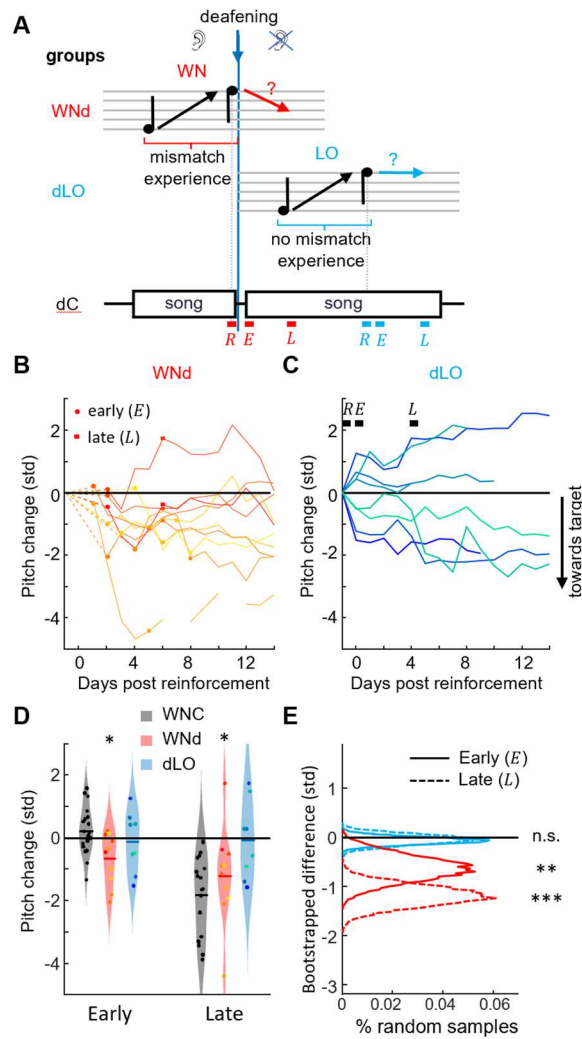
213 dLO birds were first deafened, and after they produced stable baseline song for several  
214 days, their target syllable pitch was reinforced using pitch-contingent light-off (LO)  
215 stimuli, during which the light in the sound recording chamber was briefly turned off upon  
216 high- or low-pitch syllable renditions (33). dLO birds displayed an average  $NRP$  of 112%  
217 on the 5<sup>th</sup> day since release from LO, which was significantly different from zero ( $p =$   
218  $3.7 \cdot 10^{-8}$ ,  $t_{stat} = 25.4$ ,  $N = 8$  birds, two-tailed t-test of  $H_0: NRP=0$ ) and was larger than  
219 the  $NRP$  in WNC birds on the 5<sup>th</sup> day since release ( $p = 1.3 \cdot 10^{-13}$ ,  $t_{stat} = 14.9$ ,  $df =$   
220  $24$ ,  $N = 8$  dLO and  $N = 18$  WNC birds, two-sided t-test). Thus, dLO birds were unable  
221 to recover baseline pitch, suggesting that song recovery requires undiminished sensory  
222 experience, which includes auditory feedback.

223 Deaf birds' decreased singing rate could not explain their lack of pitch recovery. Deaf  
224 birds sang less during the first 2 hours since release of reinforcement (early) than control  
225 birds:  $87 \pm 59$  motif renditions for WNd and  $410 \pm 330$  renditions for dLO compared to  
226  $616 \pm 272$  renditions for WNC birds. Also, WNd birds sang only  $4300 \pm 2300$  motif  
227 renditions between the early and late period compared to the average of  $11000 \pm$   
228  $3400$  renditions that hearing WNC birds produced in the same time period. However,  
229 despite these differences, when we inspected WNd birds' behavior 9 days after the early  
230 window, when they sang on average  $12000 \pm 6000$  renditions, their  $NRP$  was still  
231 significantly different from zero ( $NRP = 0.37$ ,  $p = 0.007$ ,  $t_{stat} = 3.47$ ,  $df = 9$ ). Thus,  
232 even after producing more practice songs than control birds, deaf birds did not recover

233 baseline pitch and so the number of songs alone cannot explain why deaf birds do not  
234 fully recover pitch. We conclude that auditory experience seems to be necessary to  
235 recover song.

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

249 Indeed, WNd birds changed their pitch significantly towards baseline already in the first 2  
250 h of their singing since release from reinforcement (relative to the pitch from the last 2 h  
251 during reinforcement). We quantified local pitch changes in terms of the  $d'$  sensitivity of  
252 signal detection theory (which is independent of shift magnitude) and found  $d' = -0.60$   
253 ( $p = 0.03$ ,  $t_{\text{stat}} = -2.19$ ,  $df = 9$ ,  $N = 10$  WNd birds, one-sided t-test of  $H_0: d' = 0$ ). A  
254 significant reversion towards pitch baseline was still evident after 4 days of practice ( $d' =$   
255  $-1.27$ ,  $p = 0.02$ ,  $t_{\text{stat}} = -2.35$ ,  $N = 10$  WNd birds, one-sided t-test, Fig. 4B, D),  
256 showing that pitch reversion in deaf birds is persistent. Because the average pitch shift in  
257 WNd birds was on the order of one standard deviation ( $d' \approx 1$ ), we conclude that without  
258 auditory experience, birds are able to perform target-directed pitch shifts of about the  
259 same magnitude as their current exploratory range (i.e., the denominator of the  $d'$   
260 measure).



**Fig. 4. Target mismatch experience is necessary for reverting 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 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 first 2 h of song after withdrawal of reinforcement (E) in std for WNd (red, B) and dLO (blue, C) birds. Curves are plotted such that pitch changes towards the target are pointing down (see Methods). (D) WNd (red) perform both early and late pitch changes in the direction of the baseline target (by about one standard deviation, \*  $p < 0.05$ , one-tailed t-test), similar to WNC (gray) and unlike dLO (blue) birds without mismatch experience. (E) Bootstrapped pitch differences between reinforced WNd (blue) and dLO (red) 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 ( $d' = -0.08$ ,  $p = 0.43$ ,  $t_{stat} = -0.18$ ,  $df = 7$ ,  $N = 8$  birds, one-tailed t-test, Fig. 4C, D). The pitch change in dLO birds was



282 indistinguishable from that in deaf controls (dC) that were not pitch reinforced (Fig. 4A,  
283 E).

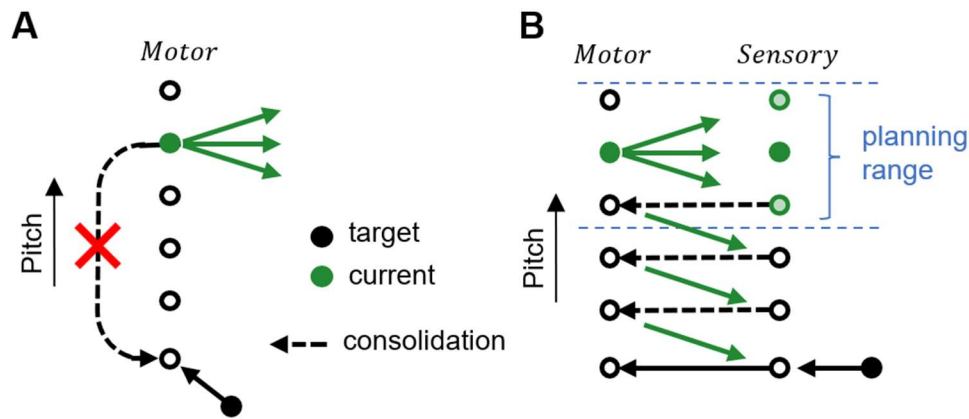
284  
285 The singing rate does not explain why deaf birds with mismatch experience partially  
286 revert their song towards baseline, unlike deaf birds without mismatch experience. WNd  
287 birds sang less during the first 2 h after reinforcement (early) than both control birds ( $p =$   
288  $2.3 \cdot 10^{-6}$ ,  $t_{\text{stat}} = 6.02$ ,  $df = 26$ ,  $N = 10$  WNd and  $N = 18$  WNC birds, two-sided t-  
289 test) and dLO birds ( $p = 0.008$ ,  $t_{\text{stat}} = 3.06$ ,  $df = 16$ ,  $N = 8$  dLO birds, two-sided t-  
290 test), unlike dLO birds that sang similar amounts as WNC birds ( $p = 0.11$ ,  $t_{\text{stat}} = 1.67$ ,  
291  $df = 24$ , two-sided t-test). If the number of songs were to determine the rate of recovery,  
292 we would have seen the opposite effect (dLO birds should recover similar amounts as  
293 WNC birds and significantly more than WNd birds). In conclusion, singing rate does not  
294 explain the difference between WNd and dLO birds.

295  
296 To discount for the effect of time elapsed since deafening and quantify the change in pitch  
297 specifically due to reinforcement, we bootstrapped the difference in  $d'$  between dLO/WNd  
298 birds and a new group of dC birds that were deafened but experienced no prior  
299 reinforcement (see methods). To discount for possible influences of circadian pitch trends,  
300 we assessed early and late pitch changes in reinforced birds and in dC birds in 2 h time  
301 windows separated by multiples of 24 h (and again flipped pitch changes in birds that  
302 were reinforced to decrease pitch, see Methods). In agreement with the findings above, we  
303 found that significant reversion towards baseline was only seen in WNd birds and very  
304 consistently so (Fig. 4E, Table S1), showing that prior experience of a target mismatch is  
305 necessary for pitch reversion independent of auditory feedback.

306  
307 We further validated our finding using a linear mixed effect model on the combined NRP  
308 data of all groups (see Methods), which confirmed our previous findings: We did not find  
309 a significant effect of the time without practice between R and E windows on the *NRP* in  
310 the E window (fixed effect  $-0.04$ ,  $p = 0.2$ ), confirming that birds do not recover without  
311 practice. Neither deafening nor muting had a significant effect by itself but the interaction  
312 between deafening and time (late) was associated with an *NRP* increase of 0.67 (fixed  
313 effect,  $p=2 \cdot 10^{-6}$ ), demonstrating that deaf birds are significantly further away from  
314 baseline ( $NRP = 0$ ) than hearing birds in late windows, thereby confirming that birds  
315 require auditory feedback to recover a distant pitch target. Importantly, we found that  
316 mismatch experience was associated with a significant fixed effect of  $-0.37$  on the *NRP*  
317 (fixed effect towards the target,  $p=0.006$ ), supporting our finding that limited vocal  
318 plasticity is possible even in the absence of auditory feedback.

319  
320 Our results thus argue for a model of song maintenance in which birds extract from target  
321 mismatch experience a plan of reducing the mismatch. Without practice and auditory  
322 experience, birds cannot reach a distant motor target (Fig. 5a). With practice and without  
323 auditory experience, they can make small changes towards a target, which we refer to as  
324 the planning range. Auditory experience allows them to consolidate the small changes  
325 such that step-by-step they can reach even a distant target (Fig. 5b).

282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328



329  
330  
331 **Fig. 5. Schematic illustrating the goal-directed planning of vocal changes.** (A) Without  
332 practice, birds cannot recover a distant motor target (black filled circle) far away  
333 from the current motor output (green filled circle). (B) Without auditory  
334 experience, birds can make motor changes (green arrows) towards a target within a  
335 small range, we refer to this range as the (overt) planning range (blue). To recover  
336 a distant target (black filled circle) beyond the planning range, birds need auditory  
337 experience (green circles under Sensory), presumably to consolidate (dashed  
338 arrows) the overt motor changes.

## 339 Discussion

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

345  
346 Our insights were gained in deaf birds and we cannot rule out that deaf birds could gain  
347 access to pitch information via somatosensory-proprioceptive sensory modalities.  
348 However, such information, even if available, cannot explain the difference between the  
349 "mismatch experience" (WNd) and the "no mismatch experience" (dLO) groups, which  
350 strengthens our claim that the pitch reversion we observe is a planned change and not  
351 merely a rigid motor response (as in simple use-dependent forgetting (5,6)). Also, it is  
352 unlikely that dLO birds' inability to recover baseline pitch is somehow due to our use of a  
353 reinforcer of a non-auditory (visual) modality, since somatosensory stimuli do not prevent  
354 reliable target pitch recovery in hearing birds (42). Thus, the overt planning ability is an  
355 active experience-dependent process.

356  
357 In our two-stage model, recovery of a developmentally learned vocal target is controlled  
358 by two hierarchical processes, a (lower) highly flexible process with limited scope ( $d' \approx$   
359 1, Fig. 4), and a dependent (higher) process enabled by experience of the former. Such  
360 motor learning based on separate processes for acquisition and retention is usually referred  
361 to as motor consolidation (4,34,35). Accordingly, the hierarchically lower process of  
362 acquisition or planning as we find is independent of immediate sensory experience, but the  
363 hierarchically higher process (consolidation of the lower process) requires experience.  
364 Perhaps then, it is the sensory experience itself that is consolidated, and therefore,  
365 consolidation of sensory experience may be a prerequisite for extensive planning.  
366

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

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

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

399 Zebra finches' ability to plan directed song changes could hinge on song memories that  
400 feed into LMAN and that could drive neurons there to produce diverse perceptual song  
401 variants. LMAN neurons are selective for the bird's own song but not the target song  
402 (21,22), which makes them well suited for executing song plans within the range of recent  
403 experience (i.e., if the song is outside recent experience, it elicits no LMAN response and  
404 so does not gain access to planning circuits). Furthermore, LMAN neurons show mirrored  
405 activity, i.e., similar activity when a zebra finch produces a vocal gesture and when it  
406 hears the same gesture played through a loudspeaker (39,40). This mirrored activity has  
407 been argued to be involved in translating an auditory target into the corresponding motor  
408 command, also known as an inverse model (41). Mirroring in LMAN was observed across  
409 the song variability generated over a period of several hours, which is about the same as  
410 the experience-dependent pitch planning limit we find. Zebra finches could thus transform  
411 a desired pitch change into the corresponding motor plan via LMAN's aligned sensory and  
412 motor representations of recent vocal output.

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

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

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

## 448 **Materials and Methods**

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

### 453 ***Subjects:***

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

### 461 ***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 (49). 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 (50), 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 (50) 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 (49). 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.

512 Two birds underwent the muting surgery directly after withdrawal of WN stimuli. To 6/8  
513 birds (4 up-shifted and 2 down-shifted), we provided the opportunity to sing without WN  
514 before the muting surgery. During on average 4h51mins (range 10 mins to 14h), these  
515 latter birds produced on average 649 song motifs (56, 100, 400, 458, 480, and 2400  
516 motifs) without WN; the example bird shown in Fig. 1C produced 56 song motifs within  
517 11 minutes during the 30 minutes it was allowed to sing without aversive reinforcement.  
518 **WN deaf (WNd)**: 10 birds were first pitch reinforced (5/10 were up-shifted and 5/10  
519 down-shifted) with WN and then they were deafened by bilateral cochlea removal. WNd  
520 birds started to sing on average  $3\pm 1$  days after deafening (range 2 to 5 days) and were  
521 recorded for at least 15 days after the deafening surgery.  
522 **Deaf LO (dLO)**: 8/10 birds from (32) were recorded after the reinforcement period and we  
523 analyzed the associated data. These birds were first deafened by bilateral cochlea removal,  
524 then they underwent pitch reinforcement with light-off (LO) stimuli that acts as an  
525 appetitive stimulus in deaf birds. The lamp in the recording chamber was switched off for  
526 100-500 ms when the pitch was either above or below a manually set threshold (daily  
527 threshold adjustment followed the same procedure as for WNm birds). 3/8 birds received  
528 LO for low-pitched syllables and 5/8 birds for high-pitched syllables. One of the birds that  
529 received LO for high-pitched syllables changed its pitch away from LO instead of towards  
530 it, thus we ended up with a balanced data set with 4/8 birds shifting pitch up and 4/8 birds  
531 shifting down. dLO birds were recorded for at least 5 days after the deafening surgery.  
532 Details of light-induced pitch shifting are described in (32).  
533 **Deaf control (dC)**: we analyzed 26 syllables from 20 birds taken (12 from (32) and 8  
534 additional ones) that were deafened and then recorded without any further manipulation.  
535 We used these birds to discount for pitch changes in WNd and dLO birds due to absence  
536 of auditory feedback, see bootstrapping.  
537 **WN duration (WNdur)**: 12 birds underwent duration reinforcement using WN, in 9 birds  
538 the targeted sound feature was syllable duration, in 2 birds the targeted feature was  
539 syllable-plus-gap duration, and in one bird the targeted feature was gap duration. In 4  
540 birds, the duration was squeezed and in 8 birds the duration was stretched. As in WNC  
541 birds, we did no further experimental manipulation after withdrawal of the WN stimulus.  
542 One bird changed its duration towards WN showed an apparent appetitive response to WN  
543 as for the one muted bird.

## 544 **Muting**

545 We muted birds by inserting a by-pass cannula into the abdominal air sac (51) as follows.  
546 Preparation of by-pass cannula: After incubation in 70% ethanol, we clogged a 7 mm long  
547 polyimide tube (diameter 1.2 mm) with sterile paper tissue. We created a suture loop  
548 around the cannula and fixed the thread to the cannula with a knot and a drop of tissue  
549 glue.  
550 Cannula implantation: We anaesthetized the birds with Isoflurane (1.5-2%) and gave a  
551 single injection of Carprofen (4 mg/kg). Subsequently, we applied local analgesic to the  
552 skin (2% lidocaine) and removed the feathers covering the right abdomen. We applied  
553 Betadine solution on the exposed skin and made a small incision using sterilized scissors.  
554 We exposed the right abdominal air sac by shifting aside the fat tissue and punctured it to  
555 create an opening. Immediately, we closed the opening by inserting the cannula and by  
556 sealing the contact region with tissue glue. With the free end of the glued thread, we made  
557 one suture to the lowest rib. We closed the wound in the skin around the cannula with  
558 tissue glue and sutures using a new thread. Finally, we applied betadine solution on the  
559 wound and lidocaine gel around the injured site. Before releasing the bird to its cage, we  
560

561 removed the clog of the cannula with forceps and verified the air flow through the  
562 cannula.

563 We returned the birds to their home cage and monitored them for signs of suffering. We  
564 administered pain killers (Meloxicam 2 mg/kg or Carprofen 2-4 mg/kg) for 2 days after  
565 the surgery.

566 On the following days, we monitored the birds continuously for singing activity. If song  
567 was detected, the cannula was inspected for clogging and cleaned. 5 birds unmuted  
568 spontaneously, they produced at most 300 songs before the bypass cannula was inspected  
569 and the clog was removed to re-mute the bird. To unclog the bypass cannulas, we used  
570 sharp forceps and sterile tissue dipped in saline. 6 of 8 birds produced quiet call-like  
571 vocalizations even on muted days on which no singing was detected.

### 572 **Deafening**

573 We bilaterally removed cochleas as described in (32).  
574

### 575 **Pitch Analysis**

576 In individual birds, we studied the dynamics of pitch recovery during the test period. In  
577 WNm birds, the test period started with unmuting, and in all other reinforced birds it  
578 started with the end of reinforcement. We analyzed songs in early (*E*) time windows  
579 defined as the first 2 h window during the test period in which the bird produced at least  
580 20 song motifs. We also assessed pitch recovery in late (*L*) windows defined exactly 4  
581 days after the *E* window. To make the measurements robust to circadian fluctuations of  
582 pitch, we compared the pitch values in early and late windows to pitch values produced in  
583 time-aligned windows during the last day of reinforcement (*R*) and during the last day of  
584 baseline (*B*).  
585

586 We used this time-of-day matched analysis to produce Fig. 1H, I, Fig. 2. C, D and Fig. 3.  
587 C, D. Exceptions where time alignment was not possible are listed in the following:

- 588 • One WNm bird started singing late on the last day of reinforcement (preventing us  
589 from time-aligning the *R* window with the *E* window), and therefore in this bird  
590 we defined *R* after the end of WN but before muting (in this bird there is more than  
591 one day of song after WN and before muting).
- 592 • In two birds (1 WNC and 1 dLO bird), we defined the *L* window one day earlier  
593 (on the 4<sup>th</sup> day, after 3 days of practice), because there was no data for these birds  
594 on the 5<sup>th</sup> day after reinforcement (our findings did not qualitatively change when  
595 we defined the *L* window on the 6<sup>th</sup> day instead of the 4<sup>th</sup>).
- 596 • One WNm bird was housed together with a female during WN reinforcement; this  
597 bird did not sing during the time-match 2-h period on the 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> day after  
598 reinforcement; therefore on those days we computed the mean pitch from all  
599 values produced on that day in Fig. 1H.

600 In early (*E*) and late (*L*) analysis windows, we computed the normalized residual pitch  
601 (*NRP*), which is the remaining fraction of pitch shift since release from WN, defined as  
602  $NRP(X) = (P_X - P_B)/(P_R - P_B)$ , where  $P_X$  is either the mean pitch in the early ( $X = E$ )  
603 or late ( $X = L$ ) window (Fig 1H, I, 2C, D, 3C, D).  $P_R$  and  $P_B$  are the mean pitches in the *R*  
604 and *B* windows, respectively. An *NRP* of 33% indicates that two-thirds of the reinforced  
605 pitch shift have been recovered and an *NRP* of 0% indicates full recovery of baseline  
606 pitch. Note that the *NRP* measure discounts for differences in the amount of initial pitch  
607 shift the birds displayed at the beginning of the test period.

608 We performed statistical testing of *NRP* to discount for this diversity in initial pitch. To  
609 test the hypothesis that WNm birds recovered their baseline pitch without practice or that

610 WNd or dLO birds recovered baseline pitch without auditory feedback, we performed a  
611 two-tailed t-test for  $NRP = 0$ .  
612 Our results were qualitatively unchanged when we changed the timing of the  $L$  window, as  
613 long as there were at least 3 days between  $E$  and  $L$  windows (because WNC birds need at  
614 least 3 days to recover their baseline pitch in the  $L$  window,  $p < 0.05$ ). Thus, giving deaf  
615 birds more time did not allow them to recover their baseline pitch. Furthermore, we also  
616 tested larger windows of 4 and 24 h duration instead of 2 h and found qualitatively similar  
617 results. We further verified that our results did not critically depend on the time-alignment  
618 by repeating the  $NRP$  tests using the last 2 h of reinforcement as the  $R$  windows. Indeed,  
619 we found that all results in Fig. 1-3 were unchanged.

620 We computed the pitch change after reinforcement (Fig. 4) as the difference in mean  
621 pitches between early ( $E$ ) or late ( $L$ ) and the last 2 h of WN/LO reinforcement  $R$  in units  
622 of sensitivity  $d' = (P_X - P_R)/S_R$ , where  $S_R$  is the standard deviation of pitch values in the  
623  $R$  window. To test the hypothesis that WNd and dLO birds are able to make targeted pitch  
624 changes towards baseline, we performed a one-tailed t-test of the hypothesis  $H_0: d' < 0$ .  
625 We used sensitivity  $d'$  relative to the last 2 h of WN/LO instead of  $NRP$  because we want  
626 to detect a pitch change, which is the realm of detection theory, i.e.  $d'$ . Furthermore, by  
627 measuring local changes in pitch relative to the last 2 h of WN/LO reinforcement, our  
628 measurements are only minimally affected by the amount of reinforcement learning that  
629 might have occurred during this 2 h time window — choosing an earlier or longer  
630 window would have blended reinforced pitch changes into our estimates. Last but not  
631 least, changes in the way in which we normalized  $d'$  values — dividing by  $S_B$ ,  
632  $\sqrt{(S_B^2 + S_R^2)/2}$ , or  $\sqrt{(S_X^2 + S_R^2)/2}$  — or using the  $NRP$  relative to the last 2 h of WN/LO  
633 did not qualitatively change the results shown in Fig. 4D.

### 634 635 **Bootstrapping**

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

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

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

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



657 distribution of average pitch change differences between WNd and dC (red) and between  
658 dLO and dC (blue) in Fig. 4E and perform bootstrap statistics.

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

### 672 **Linear Mixed Effect Model**

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

$$680 \quad NRP_{i,t} = a + b\delta_{t \in L} + c\delta_{i \in \text{WNd or } i \in \text{dLO}} + d\delta_{i \in \text{WNm}} + e\delta_{i \notin \text{dLO}} + f(t(E) - t(R))\delta_{t \in E} \\ 681 \quad \quad \quad + g\delta_{i \in \text{WNd or } i \in \text{dLO}}\delta_{i \in L} + \varepsilon_{i|\text{group}}$$

682 The Kronecker  $\delta$  of a specific group equals 1 if bird *i* belongs to that group (e.g.  $i \in \text{WN}$ )  
683 resp. if the time window *t* is either *E* or *L* (e.g.  $t \in L$ ), and it equals 0 otherwise. The  
684 terms  $\varepsilon_{i|\text{group}}$  is a random effect associated with a particular bird *i* and group. Note that  
685 the fixed effect of mismatch experience *e* is zero for dLO birds and it is one for all other  
686 birds.

687

688

### 689 **References**

- 690 1. Brecht KF, Hage SR, Gavrillov N, Nieder A. Volitional control of vocalizations in corvid  
691 songbirds. *PLoS Biol.* 2019 Aug 27;17(8):e3000375.
- 692 2. Veit L, Tian LY, Monroy Hernandez CJ, Brainard MS. Songbirds can learn flexible  
693 contextual control over syllable sequencing. *Elife.* 2021 Jun 1;10.
- 694 3. Costalunga G, Carpena CS, Selmann S, Benichov JI, Vallentin D. Wild nightingales  
695 flexibly match whistle pitch in real time. *Curr Biol.* 2023 Jul 10;
- 696 4. Brashers-Krug T, Shadmehr R, Bizzi E. Consolidation in human motor memory. *Nature.*  
697 1996 Jul 18;382(6588):252–255.

- 698 5. Galea JM, Vazquez A, Pasricha N, de Xivry J-JO, Celnik P. Dissociating the roles of the  
699 cerebellum and motor cortex during adaptive learning: the motor cortex retains what the  
700 cerebellum learns. *Cereb Cortex*. 2011 Aug;21(8):1761–1770.
- 701 6. Shmuelof L, Huang VS, Haith AM, Delnicki RJ, Mazzoni P, Krakauer JW. Overcoming  
702 motor “forgetting” through reinforcement of learned actions. *J Neurosci*. 2012 Oct  
703 17;32(42):14617–14621.
- 704 7. Krakauer JW, Ghez C, Ghilardi MF. Adaptation to visuomotor transformations:  
705 consolidation, interference, and forgetting. *J Neurosci*. 2005 Jan 12;25(2):473–478.
- 706 8. Sheahan HR, Franklin DW, Wolpert DM. Motor planning, not execution, separates motor  
707 memories. *Neuron*. 2016 Nov 23;92(4):773–779.
- 708 9. Konishi M. Birdsong: from behavior to neuron. *Annu Rev Neurosci*. 1985;8:125–170.
- 709 10. Immelmann K. Song development in the zebra finch and other estrildid finches. *Bird*  
710 *vocalizations*. 1969;
- 711 11. Park S-W, Dijkstra TMH, Sternad D. Learning to never forget-time scales and specificity of  
712 long-term memory of a motor skill. *Front Comput Neurosci*. 2013 Sep 2;7:111.
- 713 12. Smith MA, Ghazizadeh A, Shadmehr R. Interacting adaptive processes with different  
714 timescales underlie short-term motor learning. *PLoS Biol*. 2006 Jun;4(6):e179.
- 715 13. Tumer EC, Brainard MS. Performance variability enables adaptive plasticity of  
716 “crystallized” adult birdsong. *Nature*. 2007 Dec 20;450(7173):1240–1244.
- 717 14. Andalman AS, Fee MS. A basal ganglia-forebrain circuit in the songbird biases motor  
718 output to avoid vocal errors. *Proc Natl Acad Sci USA*. 2009 Jul 28;106(30):12518–12523.
- 719 15. Canopoli A, Herbst JA, Hahnloser RHR. A higher sensory brain region is involved in  
720 reversing reinforcement-induced vocal changes in a songbird. *J Neurosci*. 2014 May  
721 14;34(20):7018–7026.
- 722 16. Warren TL, Tumer EC, Charlesworth JD, Brainard MS. Mechanisms and time course of  
723 vocal learning and consolidation in the adult songbird. *J Neurophysiol*. 2011  
724 Oct;106(4):1806–1821.
- 725 17. Hoffmann LA, Saravanan V, Wood AN, He L, Sober SJ. Dopaminergic contributions to  
726 vocal learning. *J Neurosci*. 2016 Feb 17;36(7):2176–2189.
- 727 18. Aronov D, Andalman AS, Fee MS. A specialized forebrain circuit for vocal babbling in the  
728 juvenile songbird. *Science*. 2008 May 2;320(5876):630–634.
- 729 19. Nottebohm F, Stokes TM, Leonard CM. Central control of song in the canary, *Serinus*  
730 *canarius*. *J Comp Neurol*. 1976 Feb 15;165(4):457–486.
- 731 20. Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R. Neural correlates of categorical  
732 perception in learned vocal communication. *Nat Neurosci*. 2009 Feb;12(2):221–228.
- 733 21. Yazaki-Sugiyama Y, Mooney R. Sequential learning from multiple tutors and serial  
734 retuning of auditory neurons in a brain area important to birdsong learning. *J Neurophysiol*.  
735 2004 Nov;92(5):2771–2788.
- 736 22. Kojima S, Doupe AJ. Song selectivity in the pallial-basal ganglia song circuit of zebra  
737 finches raised without tutor song exposure. *J Neurophysiol*. 2007 Oct;98(4):2099–2109.
- 738 23. Yanagihara S, Yazaki-Sugiyama Y. Auditory experience-dependent cortical circuit shaping  
739 for memory formation in bird song learning. *Nat Commun*. 2016 Jun 21;7:11946.

- 740 24. Yu K, Wood WE, Theunissen FE. High-capacity auditory memory for vocal  
741 communication in a social songbird. *Sci Adv*. 2020 Nov 13;6(46).
- 742 25. Riters LV, Spool JA, Merullo DP, Hahn AH. Song practice as a rewarding form of play in  
743 songbirds. *Behav Processes*. 2019 Jun;163:91–98.
- 744 26. Kollmorgen S, Hahnloser RHR, Mante V. Nearest neighbours reveal fast and slow  
745 components of motor learning. *Nature*. 2020 Jan 8;577(7791):526–530.
- 746 27. Kojima S, Doupe AJ. Social performance reveals unexpected vocal competency in young  
747 songbirds. *Proc Natl Acad Sci USA*. 2011 Jan 25;108(4):1687–1692.
- 748 28. Elemans CPH, Rasmussen JH, Herbst CT, Düring DN, Zollinger SA, Brumm H, et al.  
749 Universal mechanisms of sound production and control in birds and mammals. *Nat*  
750 *Commun*. 2015 Nov 27;6:8978.
- 751 29. Galea JM, Mallia E, Rothwell J, Diedrichsen J. The dissociable effects of punishment and  
752 reward on motor learning. *Nat Neurosci*. 2015 Apr;18(4):597–602.
- 753 30. Suthers RA, Goller F, Wild JM. Somatosensory feedback modulates the respiratory motor  
754 program of crystallized birdsong. *Proc Natl Acad Sci USA*. 2002 Apr 16;99(8):5680–5685.
- 755 31. Yamahachi H, Zai AT, Tachibana RO, Stepien AE, Rodrigues DI, Cavé-Lopez S, et al.  
756 Undirected singing rate as a non-invasive tool for welfare monitoring in isolated male zebra  
757 finches. *PLoS One*.
- 758 32. Zai AT, Cavé-Lopez S, Rolland M, Giret N, Hahnloser RHR. Sensory substitution reveals a  
759 manipulation bias. *Nat Commun*. 2020 Nov 23;11(1):5940.
- 760 33. Zai AT, Cavé-Lopez S, Rolland M, Giret N, Hahnloser RHR. Sensory substitution reveals a  
761 manipulation bias (Data set). *ETH Research Collection*; 2020.
- 762 34. Fenn KM, Nusbaum HC, Margoliash D. Consolidation during sleep of perceptual learning  
763 of spoken language. *Nature*. 2003 Oct 9;425(6958):614–616.
- 764 35. Karni A, Sagi D. The time course of learning a visual skill. *Nature*. 1993 Sep  
765 16;365(6443):250–252.
- 766 36. Tian LY, Brainard MS. Discrete Circuits Support Generalized versus Context-Specific  
767 Vocal Learning in the Songbird. *Neuron*. 2017 Dec 6;96(5):1168–1177.e5.
- 768 37. Tachibana RO, Lee D, Kai K, Kojima S. Performance-Dependent Consolidation of Learned  
769 Vocal Changes in Adult Songbirds. *J Neurosci*. 2022 Mar 9;42(10):1974–1986.
- 770 38. Charlesworth JD, Warren TL, Brainard MS. Covert skill learning in a cortical-basal ganglia  
771 circuit. *Nature*. 2012 May 20;486(7402):251–255.
- 772 39. Hanuschkin A, Ganguli S, Hahnloser R. A Hebbian learning rule gives rise to mirror  
773 neurons and links them to control theoretic inverse models. *frontiersin.org*.
- 774 40. Oztop E, Kawato M, Arbib MA. Mirror neurons: functions, mechanisms and models.  
775 *Neurosci Lett*. 2013 Apr 12;540:43–55.
- 776 41. Giret N, Kornfeld J, Ganguli S, Hahnloser RHR. Evidence for a causal inverse model in an  
777 avian cortico-basal ganglia circuit. *Proc Natl Acad Sci USA*. 2014 Apr 22;111(16):6063–  
778 6068.
- 779 42. Sober SJ, Brainard MS. Adult birdsong is actively maintained by error correction. *Nat*  
780 *Neurosci*. 2009 Jul;12(7):927–931.

- 781 43. Zevin JD, Seidenberg MS, Bottjer SW. Limits on reacquisition of song in adult zebra  
782 finches exposed to white noise. *J Neurosci*. 2004 Jun 30;24(26):5849–5862.
- 783 44. Hough GE, Volman SF. Short-term and long-term effects of vocal distortion on song  
784 maintenance in zebra finches. *J Neurosci*. 2002 Feb 1;22(3):1177–1186.
- 785 45. Mulder T, Zijlstra S, Zijlstra W, Hochstenbach J. The role of motor imagery in learning a  
786 totally novel movement. *Exp Brain Res*. 2004 Jan;154(2):211–217.
- 787 46. Dichter BK, Breshears JD, Leonard MK, Chang EF. The control of vocal pitch in human  
788 laryngeal motor cortex. *Cell*. 2018 Jun 28;174(1):21–31.e9.
- 789 47. Belyk M, Brown S. Pitch underlies activation of the vocal system during affective  
790 vocalization. *Soc Cogn Affect Neurosci*. 2016;11(7):1078–1088.
- 791 48. Stepanek L, Doupe AJ. Activity in a cortical-basal ganglia circuit for song is required for  
792 social context-dependent vocal variability. *J Neurophysiol*. 2010 Nov;104(5):2474–2486.
- 793 49. Yamahachi H, Zai AT, Tachibana RO, Stepien AE, Rodrigues DI, Cavé-Lopez S, et al.  
794 Undirected singing rate as a non-invasive tool for welfare monitoring in isolated male zebra  
795 finches. *PLoS One*. 2020 Aug 10;15(8):e0236333.
- 796 50. Lombardino AJ, Nottebohm F. Age at deafening affects the stability of learned song in adult  
797 male zebra finches. *J Neurosci*. 2000 Jul 1;20(13):5054–5064.
- 798 51. Nilson PC, Teramitsu I, White SA. Caudal thoracic air sac cannulation in zebra finches for  
799 isoflurane anesthesia. *J Neurosci Methods*. 2005 Apr 30;143(2):107–115.

800  
801  
802

803 **Acknowledgments**

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

807  
808 **Funding:**

809 Swiss National Science Foundation (Projects 31003A\_182638 and 31003A\_156976/1)  
810 European Research Council (ERC) Advanced Grant (268911, VOTECOM)

811  
812 **Author contributions:**

813 Conceptualization: ATZ, AES, RHRH

814 Investigation: ATZ, AES, NG

815 Data Curation: ATZ, AES

816 Formal analysis : ATZ, RHRH

817 Visualization: ATZ, AES, RHRH

818 Funding acquisition: NG, RHRH

819 Supervision: NG, RHRH

820 Writing – original draft: ATZ, AES, RHRH

821 Writing – review & editing: ATZ, AES, NG, RHRH

822  
823  
824 **Competing interests:** Authors declare that they have no competing interests.

825  
826 **Data and materials availability:** Pitch and duration data that support the findings of this  
827 study together with the MATLAB scripts to reproduce the analysis and figures will be  
828 made available at the ETH Research Collection upon publication of the article. The raw  
829 data underlying the pitch measurement is not deposited due to its size but is available from  
830 the authors upon reasonable request.  
831