Article

Nearest neighbours reveal fast and slow components of motor learning

Sepp Kollmorgen[1]*, Richard H. R. Hahnloser[1] & Valerio Mante[1]*

Changes in behaviour resulting from environmental influences, development and learning[1–5] are commonly quantified on the basis of a few hand-picked features[2–4,6,7] (for example, the average pitch of acoustic vocalizations[3]), assuming discrete classes of behaviours (such as distinct vocal syllables)[2,3,8–10]. However, such methods generalize poorly across different behaviours and model systems and may miss important components of change. Here we present a more-general account of behavioural change that is based on nearest-neighbour statistics[11–13], and apply it to song development in a songbird, the zebra finch[3]. First, we introduce the concept of 'repertoire dating', whereby each rendition of a behaviour (for example, each vocalization) is assigned a repertoire time, reflecting when similar renditions were typical in the behavioural repertoire. Repertoire time isolates the components of vocal variability that are congruent with long-term changes due to vocal learning and development, and stratifies the behavioural repertoire into 'regressions', 'anticipations' and 'typical renditions'. Second, we obtain a holistic, yet low-dimensional, description of vocal change in terms of a stratified 'behavioural trajectory', revealing numerous previously unrecognized components of behavioural change on fast and slow timescales, as well as distinct patterns of overnight consolidation[1,2,4,14,15] across the behavioral repertoire. We find that diurnal changes in regressions undergo only weak consolidation, whereas anticipations and typical renditions consolidate fully. Because of its generality, our nonparametric description of how behaviour evolves relative to itself—rather than to a potentially arbitrary, experimenter-defined goal[2,3,14,16]—appears well suited for comparing learning and change across behaviours and species[17,18], as well as biological and artificial systems[5].

Zebra finches acquire complex, stereotyped vocalizations through a months-long process of sensory–motor learning[3,19–21]. During development, syllable order—that is, syntax—and the spectral structure of syllables evolve[3]. These two aspects of vocal learning may be mediated by largely independent mechanisms with distinct anatomical substrates[21,22]. Here we focus on characterizing the development of spectral structure. We began our studies by obtaining dense audio recordings of five male zebra finches between 35 and 123 days post-hatch (dph; mean ± standard deviation 73.4 ± 18.6 consecutive days of recording). Birds were isolated from other males after birth and, on average, live-tutored from around 46 to 63 dph (Extended Data Fig. 1a). Band-passed (0.35–8 kHz) audio recordings were segmented into individual vocal renditions, and represented as song spectrogram segments (Fig. 1a; 563,124–1,203,647 renditions per bird). We excluded noise and isolated calls from the analyses.

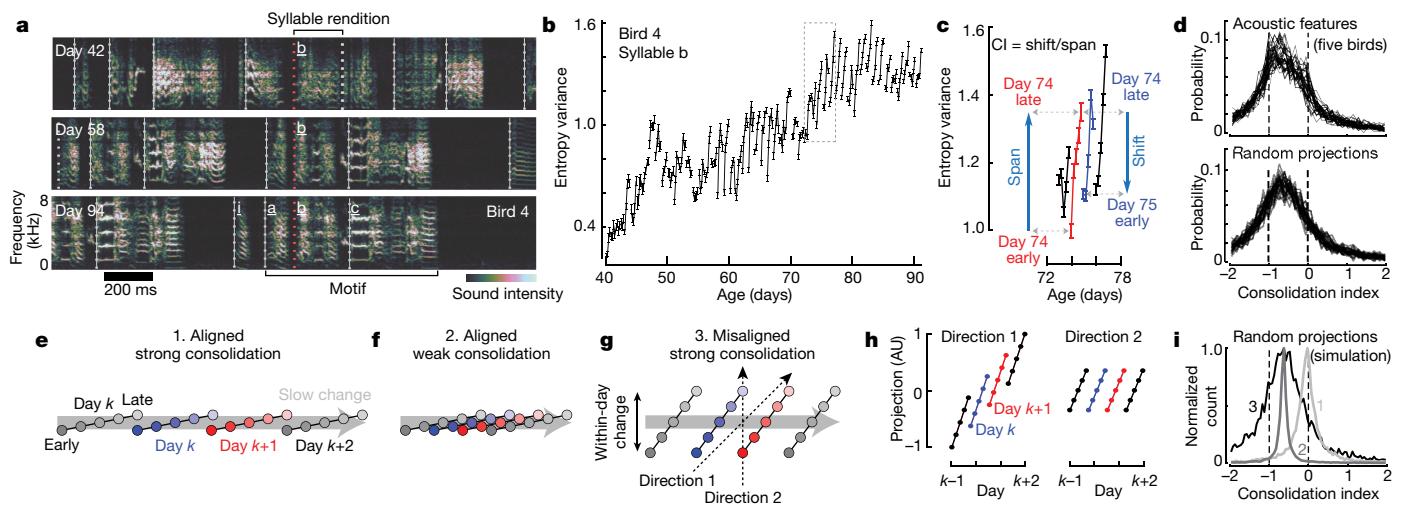### Behavioural change in single features

Vocal development is often characterized by considering changes in acoustic features such as pitch, frequency modulation[3] or entropy variance[2,14] (Fig. 1b). Such characterizations readily reveal multiple timescales of behavioural change: individual features can vary consistently within a day, display overnight discontinuities, and show drift over the duration of weeks or months (Fig. 1b, c).

We summarize the relation between change at these different timescales through a consolidation index (Fig. 1c), which measures whether within-day change in a feature ('span', Fig. 1c) is maintained or lost overnight ('shift', Fig. 1c). Weak consolidation[2,14] corresponds to a consolidation index of close to −1 (no consolidation: the shift is equal but opposite to span); strong consolidation[4,15] corresponds to an index of close to 0 (perfect consolidation: the shift is 0 days); and offline learning[4,23,24] to an index of larger than 0. Across 32 commonly used acoustic features, the consolidation indices in our data are mostly negative, indicating weak consolidation (Fig. 1d, top; median −0.67). This finding holds even for random spectral features (Fig. 1d, bottom; median −0.64) and is consistent with past accounts of song development in zebra finches[2,14].

Individual features, however, may provide an incomplete account of change in a complex behaviour such as song vocalizations. To illustrate this point, we consider three simple scenarios. In the first two (Fig. 1e, f), the change in behaviour that occurs within any given day largely mirrors, on a faster timescale, the slow change that occurs over the course

Fig. 1 | Fast and slow change in developing zebra finch vocalizations.
a, Vocalizations at three developmental stages. Dotted lines indicate syllable onsets. Crystalized song syllables (middle and bottom) fall into discrete categories (syllables i, a, b, c) and form a stereotyped 'motif', typically resembling the tutor song. **b**, Time course of one acoustic feature, entropy variance, for syllable b. **c**, Magnification of the region outlined in **b**, showing a period of within-day span (early to late, day k) and overnight shift (late day k to early day $k + 1$). The consolidation index (CI) is approximately −0.75. **d**, Histograms of consolidation indices over pairs of consecutive days, syllables and birds, for 32 acoustic features (top) and 32 random spectral projections (bottom). **e**–**g**, Three scenarios of slow developmental change (grey arrows) and fast within-day change in vocalizations. Each point represents the distribution of vocalizations from a given time and day. A larger distance between points indicates more dissimilar distributions. **h**, Linear projections of the points in **g** onto two example song features (dotted lines in **g**) for the misaligned, strong-consolidation scenario. Consolidation strength varies across directions. **i**, Consolidation indices over 10,000 random projections simulated from the three scenarios (1, 2 and 3 in **e**–**g**).

of many days or weeks. In the third scenario (Fig. 1g), within-day change is partly 'misaligned' with slow change: that is, it involves behavioural features that do not consistently change on slower timescales. Within-day change could reflect metabolic, neural or other changes that are not necessarily congruent with longer-term learning or development; the slow change reflects long-term modifications in behaviour that are typically equated with learning and development. We abstractly refer to these slow components as the direction of slow change (DiSC).

Notably, simulations of these scenarios show that negative consolidation indices for single features can result from very different time courses of development (Fig. 1h, i). Negative indices occur both when within-day and slow changes are closely aligned but daily gains along the DiSC are mostly lost overnight (weak consolidation, Fig. 1f), and when diurnal gains along the DiSC are perfectly consolidated but within-day change is substantially misaligned with slow change (Fig. 1g). The broad distributions of indices observed during song development (Fig. 1d, top), which also include strongly positive indices, seem more consistent with the misaligned scenario (Fig. 1i, histogram 3).
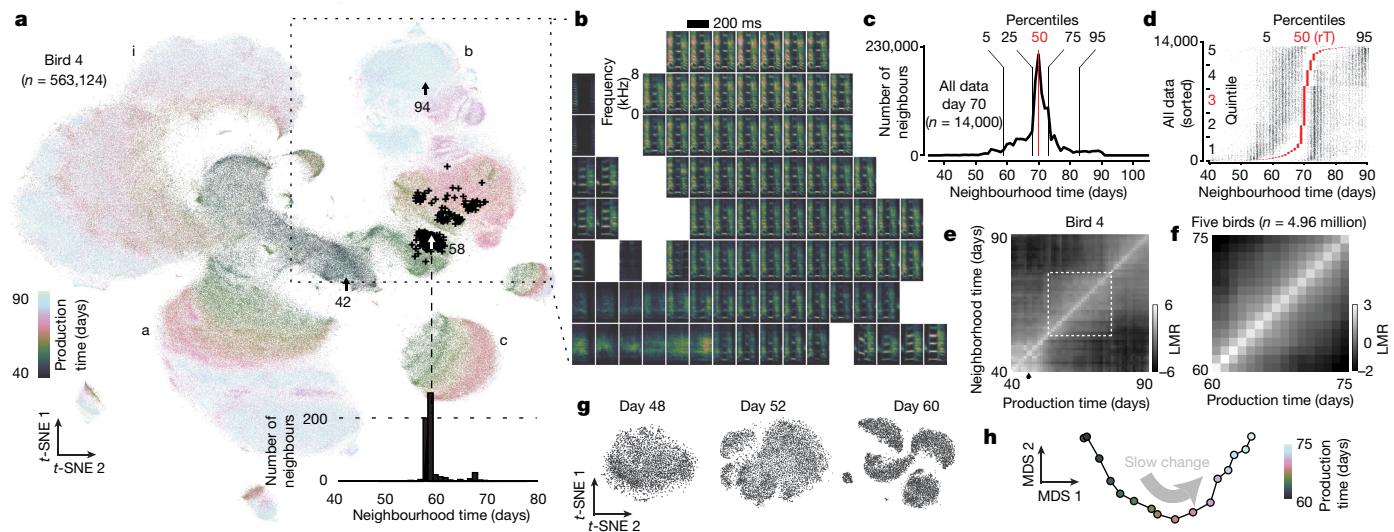
Nearest-neighbour measures of change

We developed a general characterization of change in high-dimensional behavioural data, based on nearest-neighbour statistics[12,13], that can distinguish between the scenarios in Fig. 1e–g. We initially analyse song-spectrogram segments of fixed duration aligned to syllable onset (Fig. 1a), but later extend our analysis to alternative parameterizations of the vocalization behaviour. Vocal renditions are represented as real-valued vectors $x_i \in \mathbb{R}^d$ (where i indexes renditions, and d denotes dimension), each associated with a production time, $t_i \in \mathbb{R}$ (for example, the bird's age when singing x_i). The K-neighbourhood of rendition x_i is given by those K renditions (among the set of all renditions) that are closest to x_i on the basis of some metric (for example, Euclidean distance). For small-enough values of K, different syllable types do not mix within a neighbourhood (Extended Data Fig. 1e) and neighbourhood statistics are largely independent of cluster boundaries, obviating the need for clustering renditions into syllables.

We visualize all vocalizations produced by a bird throughout development with Barnes−Hut t-distributed stochastic neighbour embedding (t-SNE)[11] (which predominantly preserves local neighbourhoods[11]). Each point in the embedding corresponds to a spectrogram segment, x_i (Fig. 1a). Different locations correspond to different vocalization types (Fig. 2b and Extended Data Fig. 2a). The embedding suggests that vocalizations change from undifferentiated subsong[3,20] (Fig. 2a, middle) to clearly differentiated syllables that fall into at least four categories (Fig. 2a, syllables a, b, c and introductory note i, as in Fig. 1a). The emergence of clustered syllables from unclustered subsong can be confirmed by standard clustering approaches (Fig. 2g and Extended Data Fig. 1c, d). Notably, the embedding does not preserve all local structure in the data, as nearest neighbours in the embedding space are not necessarily nearest neighbours in the high-dimensional data space (Fig. 2a; black crosses represent high-dimensional neighbours). We therefore quantify behavioural change directly in the high-dimensional data by analysing the composition of high-dimensional neighbourhoods[12,13] (Extended Data Fig. 2e–g).

For each data point, we refer to the production times of all data points in its K-neighbourhood as 'neighbourhood production times' (or 'neighbourhood times'; Fig. 2a, histogram). We summarize the neighbourhood times of many data points (Fig. 2d) through 'pooled neighbourhood times' (Fig. 2c) and the 'neighbourhood mixing matrix' (Fig. 2e and Extended Data Figs. 2g, 3d). Each value in the neighbourhood mixing matrix represents the similarity between behaviours from two production periods. Deviations from zero indicate that behaviours from the corresponding production periods are more similar (for values greater than 0), in terms of mixing at the level of K-neighbourhoods, or less similar (for values smaller than 0) than expected from a shuffling null hypothesis.

We use multidimensional scaling[25] on the mixing matrix to represent the similarity between behaviours from different production times as a 'behavioural trajectory' (Fig. 2h). Each point on the trajectory represents the distribution of all vocalizations produced on a given day. Pairwise distances between points represent the dissimilarity between distributions (Extended Data Fig. 2e–g). Here we focus on a 16-day

Fig. 2 | Neighbourhood mixing and repertoire dating. a, *t*-SNE of all vocalizations from the bird in Fig. 1a. Each point is a syllable rendition. Clusters (syllables i, a, b, c) emerge during development. Arrows indicate renditions of syllable b from Fig. 1a. Crosses show the 600 nearest neighbours of the rendition from day 58. Inset, histogram of production times (neighbourhood times) over the 600 nearest neighbours. **b**, Average spectrograms for different locations in the *t*-SNE visualization from **a**. **c**, Pooled neighbourhood times for day 70. Percentiles (vertical lines) quantify the extent of the behavioural repertoire on day 70. **d**, Percentiles (5th, 50th and 95th) of neighbourhood times for individual renditions from day 70 (each row represents a rendition). Rows are sorted by the 50th percentile—the repertoire time (rT, red dots). Left

and right black dots mark the 5th and 95th percentiles. A small random horizontal shift was added to each dot for visualization. **e**, Mixing matrix for all data points depicted in **a**. Each column of the matrix represents a histogram of production times, pooled over all neighbourhoods of points within a day (*x* axis), normalized by a shuffling null hypothesis (LMR, base-2 logarithm of the mixing ratio). The black arrow marks the first day of tutoring. **f**, Average mixing matrix for five birds (days 60–75). **g**, Single-day *t*-SNE, for three days (for the same bird as in Fig. 1a), illustrating the gradual emergence of clusters. **h**, Behavioural trajectory based on **f**, computed with ten-dimensional multidimensional scaling (MDS). Each point corresponds to a day. The two dimensions that capture the most variance in the trajectory are shown.

phase of gradual change midway through development (Fig. 2f). During this phase, the behavioural trajectory is structured differently on fast and slow timescales (Extended Data Fig. 3f–h). The two-dimensional projection of the trajectory that explains the maximal variance mainly reflects the direction of slow change (Fig. 1e–g, 2h).

The behavioural trajectory summarizes the progressive differentiation of vocalizations into distinct syllables, as well as simultaneous, continuous change in many spectral features of individual syllables. Notably, change is characterized through the behavioural trajectory by comparing the bird's song to itself across time, rather than to a tutor song. Thus the behavioural trajectory may also reflect innate song priors that can result in crystallized song deviating from the tutor song[26] and additional change due to other developmental processes[27].
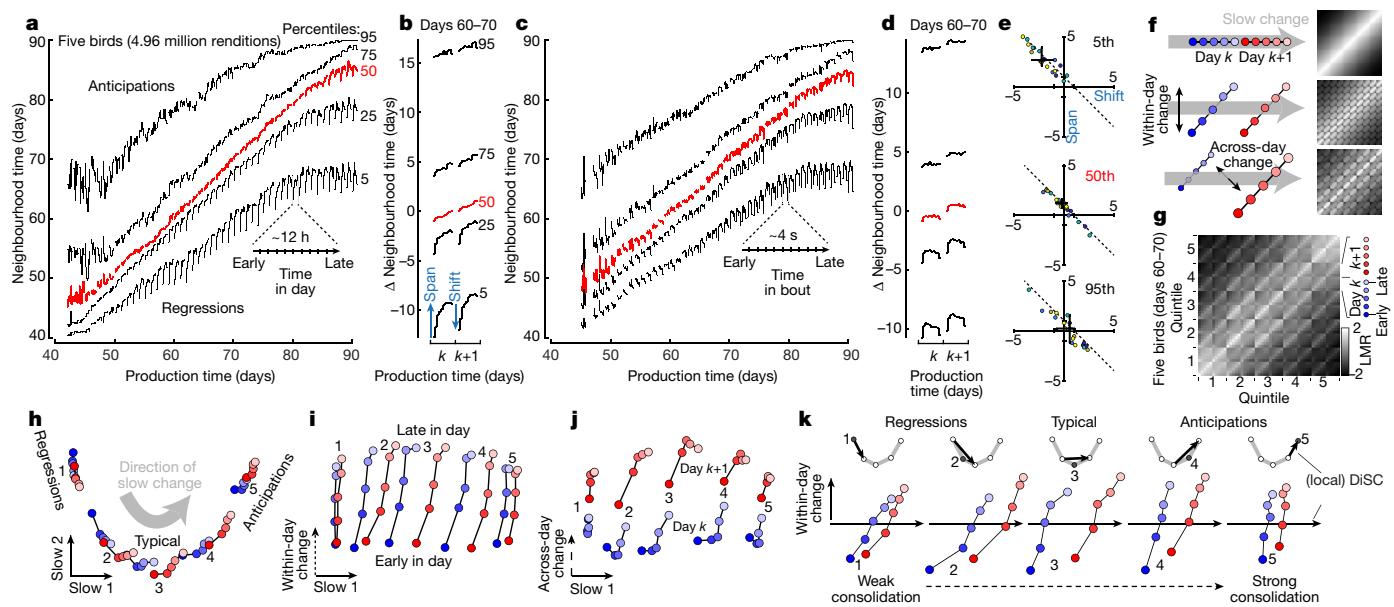
Repertoire extent and consolidation

Additional *t*-SNE visualizations of the data suggest that renditions from nearby days overlap considerably, whereby changes occurring within a day partly mimic the slow change across days (Extended Data Fig. 2b, c). We quantify this apparent spread along the DiSC—reflecting different degrees of behavioural 'maturity'—through neighbourhood times (Fig. 2d). We refer to behavioural renditions that predominantly have neighbours produced in the future as 'anticipations', and to renditions that predominantly have neighbours that were produced in the past as 'regressions' (Extended Data Fig. 3b). By contrast, renditions that are 'typical' for a given developmental stage mostly have neighbours produced on the same or nearby days. We denote the median neighbourhood time as the 'repertoire time' of a rendition. The repertoire time effectively places each rendition along the DiSC (Fig. 2d, *x* axis): that is, it dates it with respect to the progression of vocal development ('repertoire dating'). A broad distribution of repertoire times across all renditions in a day (Fig. 2d) suggests considerable behavioural variability along the DiSC; the most extreme regressions are backdated

more than ten days into the past, and the most extreme anticipations are post-dated more than ten days into the future.

To quantify behavioural change on the timescale of hours, we subdivide each day into ten consecutive periods, and compute pooled neighbourhood times separately for each period. The percentiles of the pooled neighbourhood times chart the evolution of behaviour within and across days throughout development (Fig. 3a). Each repertoire-dating percentile is akin to a learning curve for a part of the behavioural repertoire (for example, typical renditions are described by the 50th percentile, and extreme anticipations by the 95th). The evolution of each percentile captures the progress along the DiSC (Fig. 3a, *y* axis) over time (Fig. 3a, *x* axis). We validated this characterization of behavioural change on simulated behaviour that mimicked vocal development (Extended Data Fig. 4a–d).

The repertoire-dating percentiles reveal that typical renditions move gradually along the DiSC throughout the day, and that changes along the DiSC acquired during the day are, on average, fully consolidated overnight (Fig. 3a, b, red). Anticipations undergo a similar or smaller degree of within-day change (Fig. 3a, b, 75th and 95th percentiles), whereas regressions move by a larger distance within each day, but this change is only weakly consolidated overnight (Fig. 3a, b, 5th and 25th percentiles; Fig. 3e). The most 'immature' renditions thus improve markedly throughout a day—more than typical renditions or anticipations—but these improvements are mostly lost overnight. This pattern of change seems to be characteristic of development, as it is absent in adults (Extended Data Figs. 5, 6).

Movement along the DiSC also occurs on timescales that are faster than hours, namely within bouts of singing—that is, groups of vocalizations that are preceded and followed by a pause (average bout duration 3.81 ± 0.83 s across birds). We subdivide each bout into ten consecutive periods, compute pooled neighbourhood times for each period (over all bouts in a day), and track change through the corresponding percentiles (Fig. 3c, d). Within bouts, large changes along the DiSC occur at

Fig. 3 | Multiple components of behavioural change during sensory–motor learning. a, Average repertoire dating percentiles (for five birds) describing within and across-day changes along the DiSC. For each production day and period, five percentiles of the pooled neighbourhood times (Fig. 2c) are arranged vertically (lines). **b**, Average of data from **a** across days 60–70, expressed relative to the average 50th percentile. **c**, Within-bout changes. As for **a**, but based on production day and period in a singing bout. **d**, As for **b**, but averaged across data from **c**. **e**, Span and shift for the 5th, 50th and 95th percentiles (blue arrows in **b**, analogous to Fig. 1c) averaged over days 50–80, separately for syllables (points) and birds (colours). Black lines indicate medians and 95% bootstrapped confidence intervals over all points.

f, Simulated stratified mixing matrices (right) for three models (left) of the alignment of within-day and across-day change with the DiSC. **g**, Average measured stratified mixing matrices (five birds, days 60–70). **h–j**, Stratified behavioural trajectory based on **g**. Different two-dimensional projections reveal the DiSC (**h**), as well as within-day (**i**) and across-day (**j**) change not aligned with the DiSC (labels 1–5 represent different strata). The full ten-dimensional trajectories faithfully reproduce the structure of the stratified mixing matrices (MDS stress = 0.016); the depicted four-dimensional subspace captures 81% of the ten-dimensional variance. **k**, Separate projections for each stratum onto the local DiSC (black arrows in upper diagrams; points represent strata from **h**).

the regressive tail of the behavioural repertoire: vocalizations are most regressive at the onset and offset of bouts (Fig. 3c, d, 5th percentile). Similar, albeit weaker changes occur for typical renditions (Fig. 3c, d, red). The same apparent changes in song maturity are observed when short and long bouts (durations 2.30 ± 0.54 s versus 6.28 ± 1.73 s) are considered separately. Song maturity thus decreases at the end of a bout, not after a fixed time into the bout (Extended Data Fig. 5a–c).

Misaligned behavioural components

The repertoire time reveals within-day and within-bout changes that mirror, on a faster timescale, changes that also occur over many days (see Supplementary Methods). As above (Fig. 1), we refer to such components of change as being aligned with the DiSC, and to components that are not reflected in the repertoire time as being misaligned.

We identify both aligned and misaligned components of change through the 'stratified mixing matrix', which combines a neighbourhood-mixing matrix (for example, Fig. 2f) with repertoire dating. Each day's behavioural repertoire is binned into five consecutive production periods. Within each period, the behavioural repertoire is subdivided into five strata on the basis of repertoire time (Fig. 2d, quintiles). All renditions from a day thus fall into 5 × 5 = 25 bins. The stratified mixing matrix measures similarity between 50 bins that combine the data from two adjacent days (Fig. 3g). We compare the measured stratified mixing matrix with simulations that differ with respect to how within-day change and change across adjacent days align with the DiSC (Fig. 3f and Extended Data Fig. 4e–j). In model 1, development is one-dimensional and therefore aligned with the DiSC (Fig. 3f, top; similar to Fig. 1e). In model 2, within-day change involves a component that is not aligned with the DiSC (Fig. 3f, middle; similar to Fig. 1g). In model 3, adjacent days are separated not only along the DiSC, but also along a direction

orthogonal to both the DiSC and the direction of within-day change (Fig. 3f, bottom, across-day change). Prominent 'stripes' along every other diagonal in the measured mixing matrix (Fig. 3g) indicate a larger similarity between renditions from the same day than between renditions from adjacent days, as predicted by model 3, suggesting that several misaligned components contribute to change at fast timescales.

From the stratified mixing matrix, we infer stratified behavioural trajectories. The two-dimensional projection that captures most of the variance due to strata (Fig. 3h) resembles Fig. 2h and reflects the DiSC. Consistent with repertoire dating, behavioural change along the DiSC between adjacent days (Fig. 3h, blue versus red for each stratum) is small compared with the spread of the behaviour for one day along the DiSC (for example, blue points, strata 1–5). For each stratum, however, much of the change that occurs within a day is misaligned with the DiSC (Fig. 3i, k; early versus late separated along the orthogonal dimension of within-day change). Yet another misaligned component is necessary to appropriately capture change across adjacent days (Fig. 3j). These properties of aligned and misaligned components are replicated by a linear analysis based on spectral features that are chosen to capture change at specific timescales (Extended Data Figs. 7, 8), and are robust to how song is parameterized and segmented, and to how nearest neighbours are defined (Extended Data Figs. 9, 10).

Discussion

Our analysis of high-dimensional vocalizations reveals that vocal learning and development do not reflect an underlying one-dimensional process. Single behavioural features in isolation therefore provide an incomplete account of behavioural change during development and learning. The weak consolidation observed here (Fig. 1d) and elsewhere[2,14] at the level of single features appears to reflect prominent

Article

misaligned components of within-day change, rather than weak consolidation along the DiSC (Fig. 1h, i). Strong overnight consolidation along the DiSC across much of the behavioural repertoire (Fig. 3a, b) seems consistent with consolidation patterns observed for skilled motor learning in humans[23,24,28] and of motor adaptation in humans[1,18] and birds[4].

Our characterization of behaviour on the basis of nearest-neighbour statistics can be applied when no accurate parametric model of the behaviour is known, as is the case at present for most natural, complex behaviours. The approach is largely complementary to methods that rely on clustering behaviour into distinct categories[2,3,10,29]. Forgoing an explicit clustering of the data can be advantageous, because assuming the existence of clusters can be an unwarranted approximation[30] and may impede the characterization of behaviour that appears not to be clustered (such as juvenile zebra finch song; Extended Data Fig. 1); moreover, determining correct cluster boundaries is in general an ill-defined problem[30]. Notably, our analyses require only an indicator function that selects nearest neighbours (based here on a 'locally meaningful' distance metric)—a much weaker requirement than a globally valid distance metric or the existence of a low-dimensional feature space that globally maps behavioural space[11]. These properties make repertoire dating applicable to almost any behaviour and other high-dimensional datasets, including data that are characterized by 'labels' other than production time. Repertoire dating may thus provide a general account of learning and change that is amenable to comparisons between different behaviours and model systems, including different species[17] and artificial systems[5].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1892-x.

1. Brashers-Krug, T., Shadmehr, R. & Bizzi, E. Consolidation in human motor memory. *Nature* **382**, 252–255 (1996).
2. Derégnaucourt, S., Mitra, P. P., Fehér, O., Pytte, C. & Tchernichovski, O. How sleep affects the developmental learning of bird song. *Nature* **433**, 710–716 (2005).
3. Tchernichovski, O., Mitra, P. P., Lints, T. & Nottebohm, F. Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* **291**, 2564–2569 (2001).
4. Andalman, A. S. & Fee, M. S. A basal ganglia–forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc. Natl Acad. Sci. USA* **106**, 12518–12523 (2009).
5. Arulkumaran, K., Deisenroth, M. P., Brundage, M. & Bharath, A. A. Deep reinforcement learning: a brief survey. *IEEE Signal Process. Mag.* **34**, 26–38 (2017).
6. Ingram, J. N., Flanagan, J. R. & Wolpert, D. M. Context-dependent decay of motor memories during skill acquisition. *Curr. Biol.* **23**, 1107–1112 (2013).
7. Klaus, A. et al. The spatiotemporal organization of the striatum encodes action space. *Neuron* **95**, 1171–1180 (2017).
8. Han, S., Taralova, E., Dupre, C. & Yuste, R. Comprehensive machine learning analysis of *Hydra* behavior reveals a stable basal behavioral repertoire. *eLife* **7**, e32605 (2018).
9. Egnor, S. E. R. & Branson, K. Computational analysis of behavior. *Annu. Rev. Neurosci.* **39**, 217–236 (2016).
10. Wiltschko, A. B. et al. Mapping sub-second structure in mouse behavior. *Neuron* **88**, 1121–1135 (2015).
11. van der Maaten, L. Accelerating *t*-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).
12. Chen, H. & Friedman, J. J. H. A new graph-based two-sample test for multivariate and object data. *J. Am. Stat. Assoc.* **1459**, 1–41 (2016).
13. Hawks, M. *Graph-Theoretic Statistical Methods for Detecting and Localizing Distributional Change in Multivariate Data*. PhD Thesis, Naval Postgraduate School, Monterey, California (2015).
14. Shank, S. S. & Margoliash, D. Sleep and sensorimotor integration during early vocal learning in a songbird. *Nature* **458**, 73–77 (2009).
15. Fenn, K. M., Nusbaum, H. C. & Margoliash, D. Consolidation during sleep of perceptual learning of spoken language. *Nature* **425**, 614–616 (2003).
16. Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B. & Mitra, P. P. A procedure for an automated measurement of song similarity. *Anim. Behav.* **59**, 1167–1176 (2000).
17. Anderson, D. J. J. & Perona, P. Toward a science of computational ethology. *Neuron* **84**, 18–31 (2014).
18. Krakauer, J. W. & Shadmehr, R. Consolidation of motor memory. *Trends Neurosci.* **29**, 58–64 (2006).
19. Brainard, M. S. & Doupe, A. J. What songbirds teach us about learning. *Nature* **417**, 351–358 (2002).
20. Catchpole, C. K. & Slater, P. J. B. *Bird Song: Biological Themes and Variations* (Cambridge Univ. Press, 2003).
21. Dhawale, A. K., Smith, M. A. & Ölveczky, B. P. The role of variability in motor learning. *Annu. Rev. Neurosci.* **40**, 479–498 (2017).
22. Lipkind, D. et al. Songbirds work around computational complexity by learning song vocabulary independently of sequence. *Nat. Commun.* **8**, 1247 (2017).
23. Korman, M. et al. Daytime sleep condenses the time course of motor memory consolidation. *Nat. Neurosci.* **10**, 1206–1213 (2007).
24. Fischer, S., Hallschmid, M., Elsner, A. L. & Born, J. Sleep forms memory for finger skills. *Proc. Natl Acad. Sci. USA* **99**, 11987–11991 (2002).
25. Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27 (1964).
26. Fehér, O., Wang, H., Saar, S., Mitra, P. P. & Tchernichovski, O. De novo establishment of wild-type song culture in the zebra finch. *Nature* **459**, 564–568 (2009).
27. Adam, I. & Elemans, C. P. H. Vocal motor performance in birdsong requires brain–body interaction. *eNeuro* **6**, ENEURO.0053-19.2019 (2019).
28. Walker, M. P., Brakefield, T., Hobson, J. A. & Stickgold, R. Dissociable stages of human memory consolidation and reconsolidation. *Nature* **425**, 616–620 (2003).
29. Vogelstein, J. T. et al. Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science* **344**, 386–392 (2014).
30. Fahad, A. et al. A survey of clustering algorithms for big data: taxonomy and empirical analysis. *IEEE Trans. Emerg. Top. Comput.* **2**, 267–279 (2014).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Code availability

We provide source code for our nearest-neighbour-based analyses (https://github.com/skollmor/repertoireDating). We also provide a data-browser/data analysis graphical user interface that we used to perform some analyses and to explore and visualize data (https://github.com/skollmor/dspace).

Author contributions S.K. conceived of the approach. S.K. and V.M. performed analyses. S.K., V.M. and R.H.R.H. wrote the paper. R.H.R.H. conceived and supervised collection of the behavioural data.
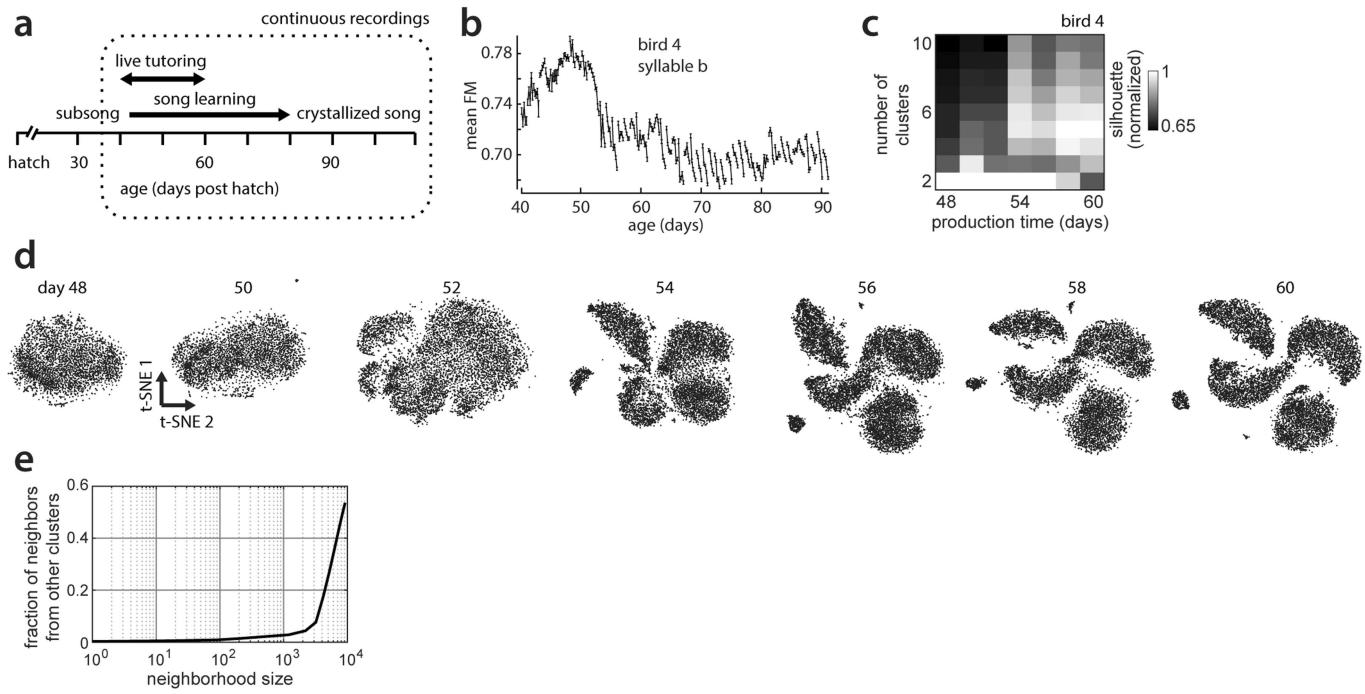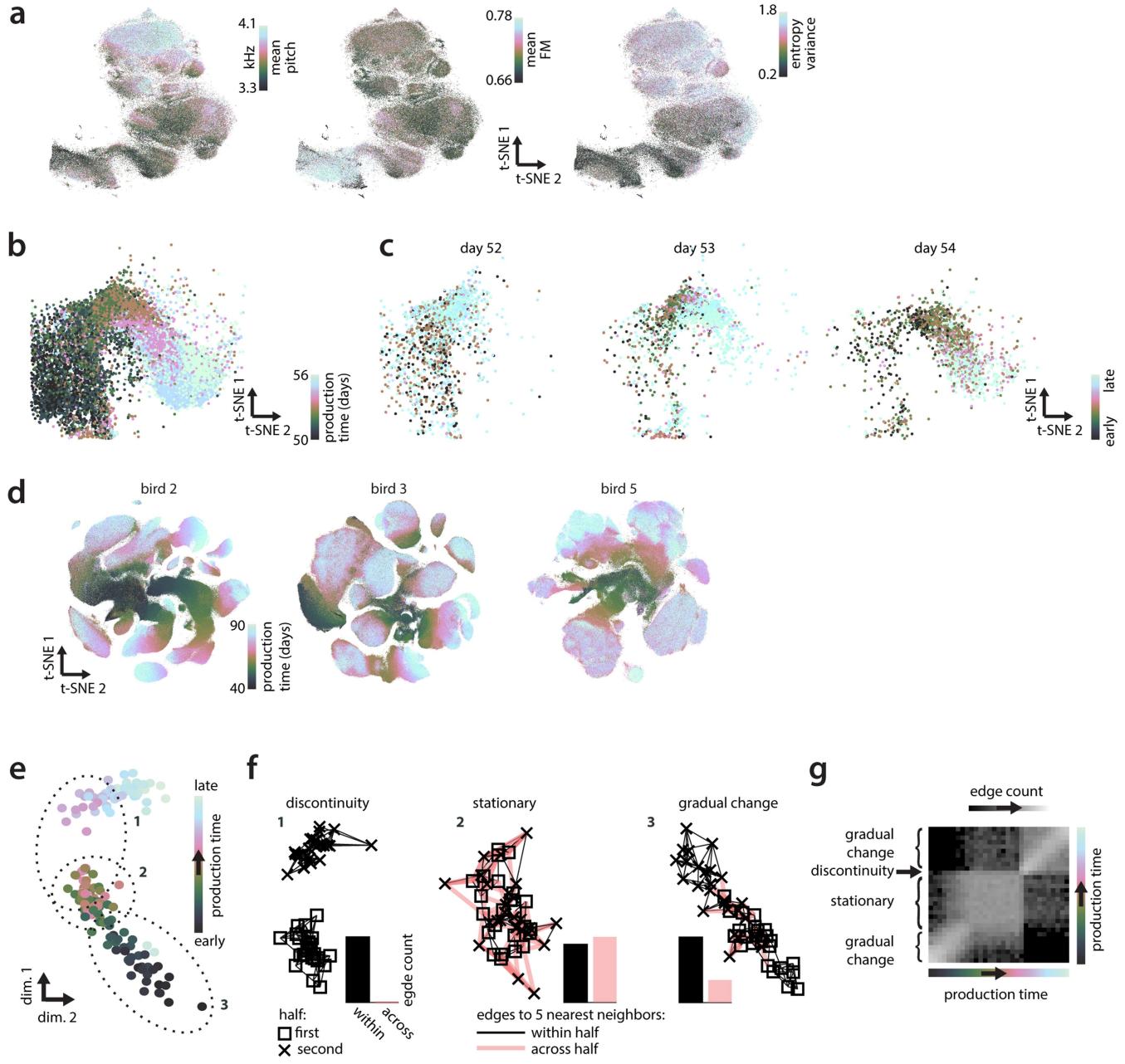
Article

Extended Data Fig. 1 | Clustering of juvenile and adult zebra finch song.
a, Vocal development in male zebra finches. Tutoring by an adult male started at around day 46 (post-hatch) and lasted 10–20 days. **b**, Time course of the acoustic feature frequency modulation (FM), for syllable b in the example bird (compare with Fig. 1b). **c**, Normalized mean silhouette values for 2–10 clusters for vocalizations from the seven days shown in **d**. High values indicate evidence for the respective cluster count. Normalized mean silhouette coefficients are based on 20 repetitions of k-means clustering of random subsets of 1,000 68-ms onset-aligned spectrogram segments from a single day (as in **d**), projected onto the first five principal components. **d**, t-SNE visualizations of vocalizations produced on a given day post-hatch for the example bird (bird 4, the same bird as in Fig. 1a–c). A separate embedding was computed for each day, and the embedding's initial condition was based on the previous day. Note

the gradual emergence of clusters, each corresponding to a distinct syllable type (for example, syllables i, a, b, c in Fig. 2a). **e**, Average fraction of neighbours from a different cluster, as a function of neighbourhood size. These data are analogous to those from **c**, **d** but for vocalizations from day 90 (12,854 data points), when clusters are fully developed. For a wide range of neighbourhood sizes, the neighbours of a data point mostly belong to the same cluster or syllable type. For a neighbourhood size of 100, the average fraction of out-of-cluster neighbours from the same day is 0.0089. Thus, for an appropriately chosen neighbourhood size, nearest-neighbour methods respect clustering structure in the data by construction, and sidestep having to explicitly identify clusters in the data. In most analyses, we computed nearest neighbours for data from all days, meaning that clustering structure is respected even for neighbourhood sizes that are slightly larger than those suggested in **e**.
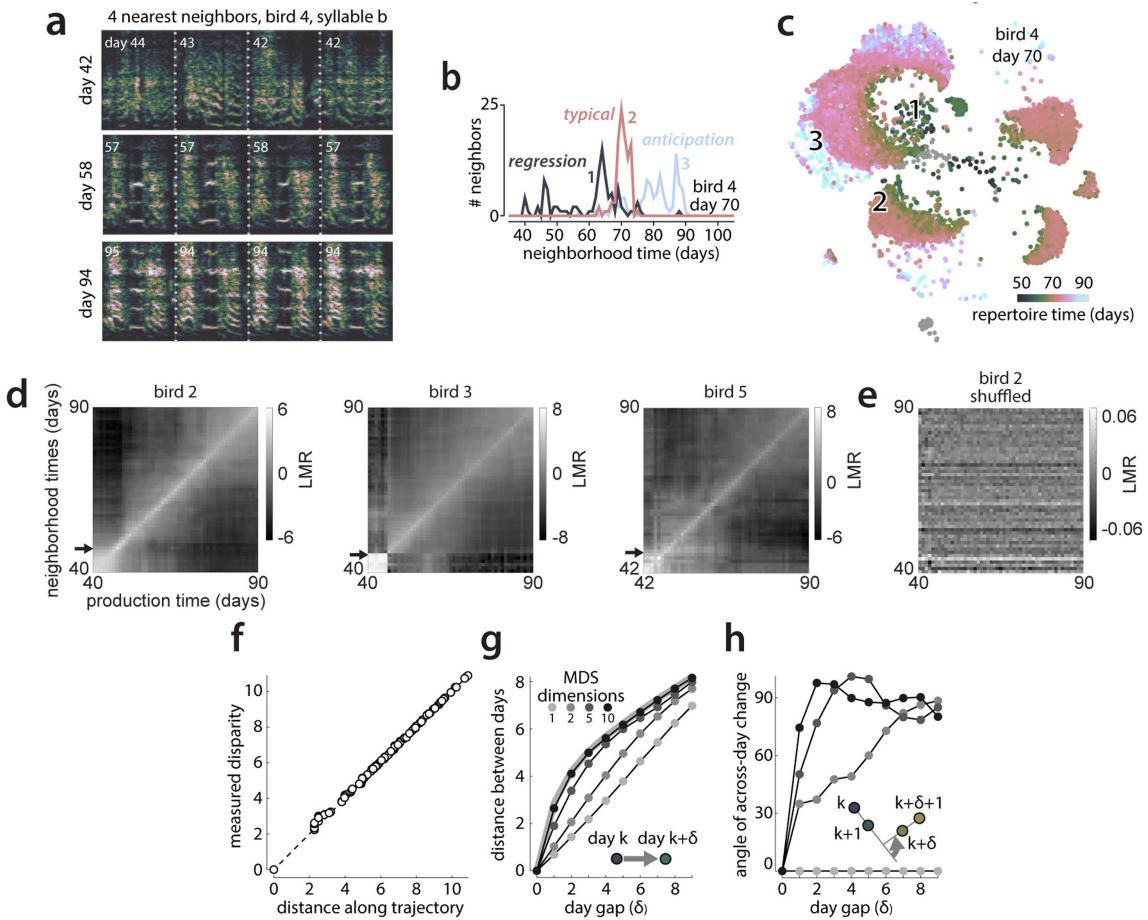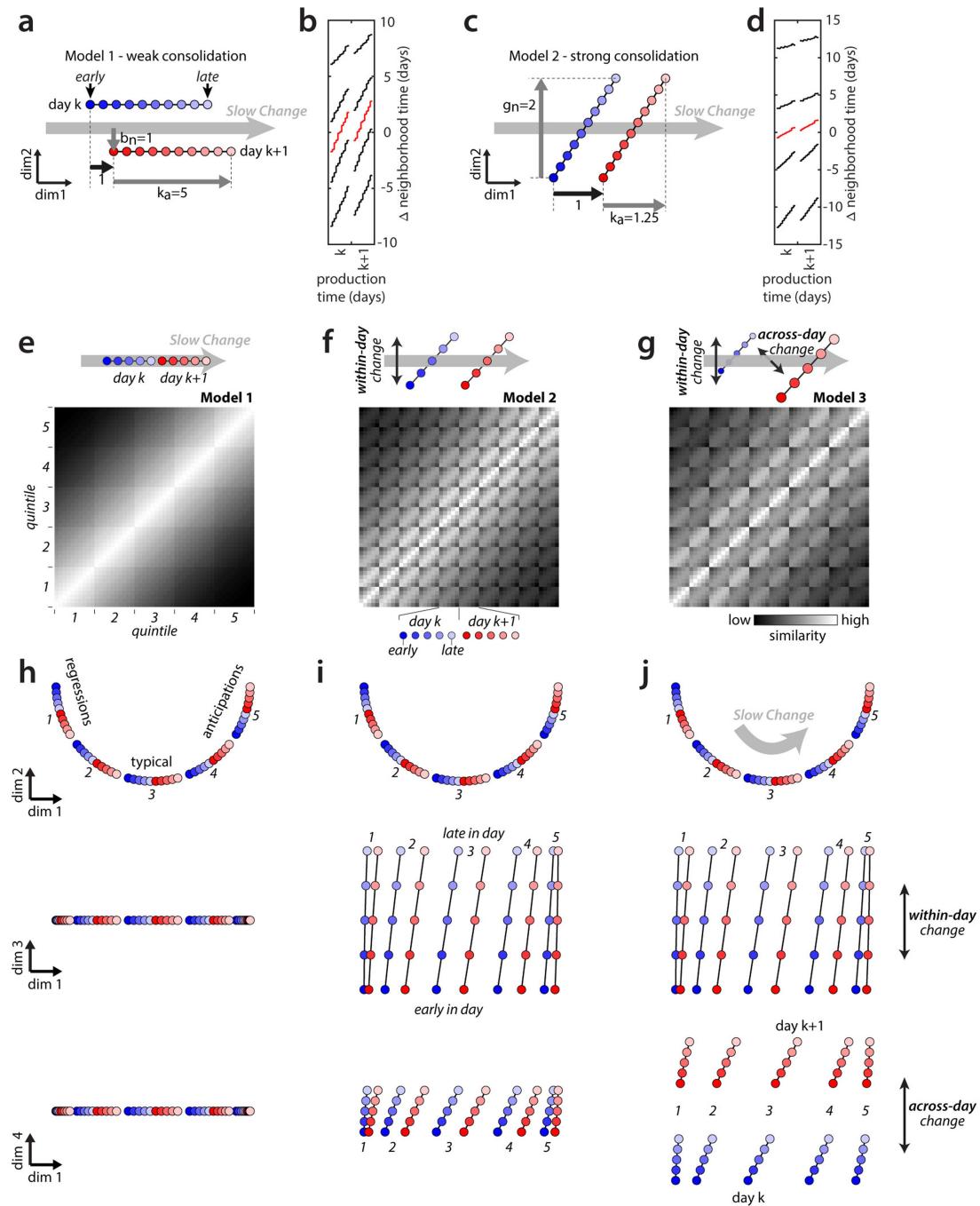
Extended Data Fig. 2 | See next page for caption.

Article

Extended Data Fig. 2 | Properties of large-scale embeddings. a, Three auditory features computed on renditions of syllable b. This panel uses the same embedding as in Fig. 2a, but with different colours. **b**, Across-day change in vocalizations. This is a magnified cutout from the bottom left region of the dashed outline of Fig. 2a. The colours differ from Fig. 2a, and points from days 50–56 only are shown. **c**, Within-day change in vocalizations. Points from **b** are shown separately for three individual days and coloured according to production time within the day (early to late). Vocalizations change within a day: early vocalizations (dark green) are more similar to vocalizations from previous days (dark green points in **b**); late vocalizations (light blue) are more similar to vocalizations from future days (light blue in **b**). **d**, *t*-SNE visualizations for dense recordings from three birds (analogous to Fig. 2a). **e**–**g**, Illustration of a fictitious behaviour that undergoes distinct phases of abrupt change, no change and gradual change, and the identification of these phases on the basis of nearest-neighbour graphs. **e**, A low-dimensional representation of the behaviour. Each point corresponds to a behavioural rendition (for example, a syllable rendition) and is coloured according to production time. Similar renditions (for example, syllable renditions with similar spectrograms) appear near each other in this representation. The dotted ellipses mark three subsets of points corresponding to: (1) a phase of abrupt change; (2) a phase of no change; and (3) a phase of gradual change. **f**, Nearest-neighbour graphs for the three subsets of points in **e**. Points are replotted from **e** with different symbols, indicating whether their production times fall within the first half (squares) or second half (crosses) of the corresponding subset. Edges connect each point to its five nearest neighbours. The edge colour marks neighbouring pairs of points falling into the same (black) or different (red) halves. Relative counts of within- and across-half edges differ according to the nature of the underlying behavioural change (histograms of edge counts). If an abrupt change in behaviour occurs between the first and second half, nearest neighbours of points in one half will all be points from the same half, and none from the other half (discontinuity). When behaviour is stationary, the neighbourhoods are maximally mixed: that is, every point has about an equal number of neighbours from the two halves. Phases of gradual change result in intermediate levels of mixing. **g**, Mixing matrix for the simulated data in **e**, analogous to Fig. 2e. Each location in the matrix corresponds to a pair of production times. Strong mixing (white) indicates a large number of nearest-neighbour edges across the two corresponding production times (as in **f**; stationary) and thus similar behaviour at the two times. Weak mixing (black) indicates a small number of such edges (as in **f**; discontinuity), and thus dissimilar behaviour. Note that such statistics on the composition of local neighbourhoods can be computed for any kind of behaviour and are invariant with respect to transformations of the data that preserve nearest neighbours, such as scaling, translation and rotation. These properties make nearest-neighbour approaches highly general.

Extended Data Fig. 3 | Repertoire dating and the direction of slow change. a, The four nearest neighbours for example vocalizations (bird 4, syllable b, from Fig. 1a). Production times of nearest neighbours (numbers) need not equal that of the corresponding example rendition. **b**, Neighbourhood production times for three renditions from day 70 (analogous to Fig. 2a, inset). Rendition 2 is 'typical' for day 70 (most neighbours lie in the same or adjacent days); renditions 1 and 3 are a 'regression' and an 'anticipation' (with neighbours predominantly produced in the past or future, respectively). **c**, All renditions from day 70 (a subset of the points in Fig. 2a). Colours correspond to repertoire time (50th percentile in Fig. 2d). Anticipations (repertoire times greater than 70) and regressions (repertoire times less than 70) occur at locations corresponding to vocalizations typical of later and earlier development (compare with Fig. 2a). Numbers 1–3 mark the approximate locations of the example renditions in **b**. **d**, Mixing matrices for additional birds (analogous to Fig. 2e, using the same birds as in Extended Data Fig. 2d). Bird 3 produced only a very few vocalizations (mostly calls) before tutoring onset (black arrows). The mixing matrices consistently show a period of gradual change starting after tutor onset and lasting several weeks. This gradual change typically slows down (resulting in larger mixing values far from the diagonal) at the end of the developmental period considered here (day 90 post-hatch; later periods are in Extended Data Fig. 6). Grey values correspond to the base-2 logarithm of the mixing ratio (LMR), that is, histograms over the pooled neighbourhood times (Fig. 2c) normalized by a null hypothesis obtained from a random distribution of production times (see Supplementary Methods). For example, an LMR value of 5 implies that renditions from the corresponding pair of production times are $2^5 = 32$ times more mixed at the level of local neighbourhoods than would be expected by chance (that is, there is a random distribution of production times across renditions). **e**, As in **d**, bird 2, but after shuffling production times among all data points. Effects under this null hypothesis are small (the maximal observed mixing ratio is $2^{0.06}$ or approximately 1.042). Similar, small effects under the null hypothesis are obtained for the other mixing matrices discussed throughout the text. **f–h**, Properties of the behavioural trajectory inferred from the mixing matrix in Fig. 2f. **f**, Pairwise distances between points along the inferred behavioural trajectory (x axis), plotted against measured disparities

(y axis). Disparities are obtained by rescaling and inverting the similarities in Fig. 2f (see Supplementary Methods). The points on the trajectory are inferred with ten-dimensional non-metric MDS on the measured disparities. Importantly, the pairwise distances between inferred points faithfully represent the corresponding, measured disparities (all points lie close to the diagonal; MDS stress = 0.0002). **g, h**, Structure of low-dimensional projections of the behavioural trajectory. We applied principle-component analysis to the ten-dimensional arrangement of points inferred with MDS and retained an increasing number of dimensions (number of dimensions indicated by greyscale). For example, the projection onto the first two principle components is shown in Fig. 2h (MDS dimension 2 in **g, h**). The first two principle components explain 75% of the variance in the full ten-dimensional trajectory. **g**, Measured (true) disparity (thick grey curve) and distances along the inferred trajectories (points and thin curves) as a function of the day gap (δ) between points. For any choice of projection dimensionality and δ, we computed the Euclidean distances between any two points separated by δ and averaged across pairs of points. The measured (true) disparities increase rapidly between subsequent and nearby days, but only slowly between far apart days (thick grey curve). Low-dimensional projections of the trajectory (for example, MDS dimension 2) underestimate the initial increase in disparities. **h**, Angle between the reconstructed direction of across-day change for inferred behavioural trajectories, as a function of the day gap between points. Same conventions and legend as in **g**. For the one- and two-dimensional trajectories, the direction of across-day change varies little or not at all from day to day (see inset; the arrow indicates the angle of across-day change). On the other hand, the direction of across-day change along the full, ten-dimensional behavioural trajectory is almost orthogonal for subsequent days. Data shown in **g, h** suggest that the full behavioural trajectory is more 'rugged' than indicated by the two-dimensional projection in Fig. 2h. This structure is consistent with the finding that across-day change includes a large component that is orthogonal to the directions of slow change and of within-day change (Fig. 3j). Note that **a** shows 200-ms spectrogram segments, whereas **b–h** are based on 68-ms segments (as are most of the analyses).
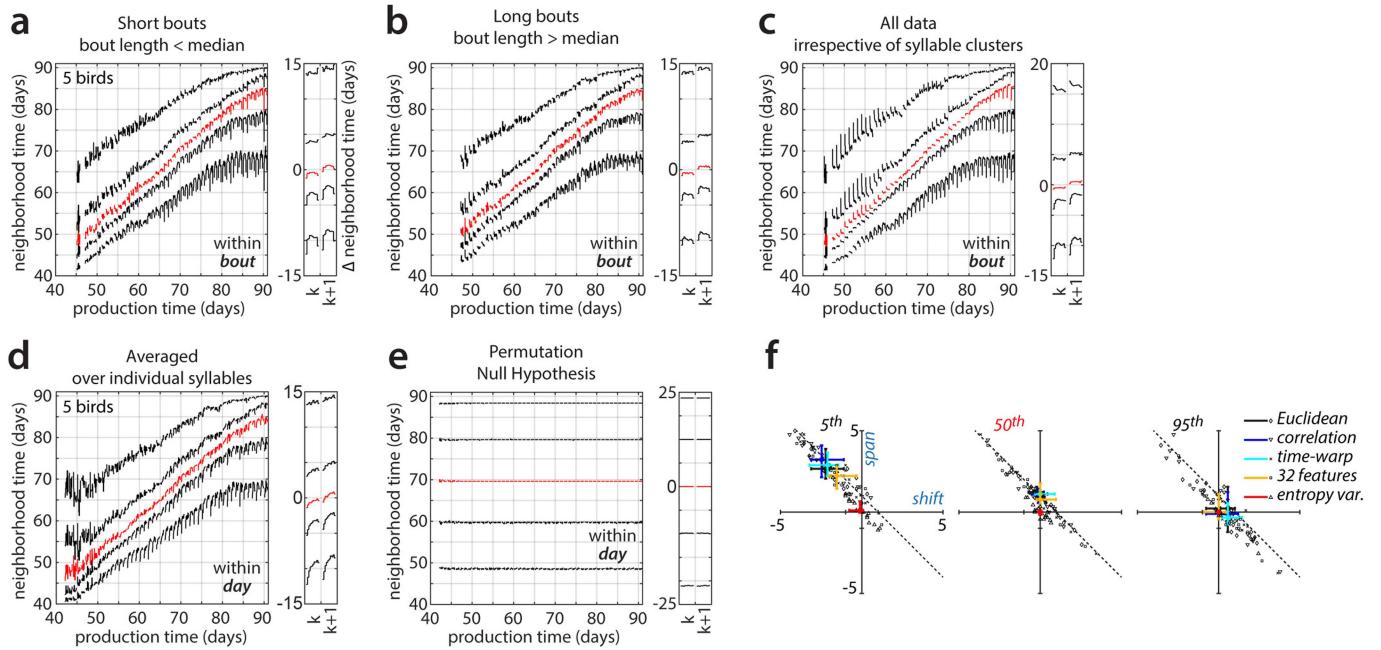
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Models of the alignment between components of change. a–d, Validation of repertoire dating. We simulated individual behavioural renditions as points in a high-dimensional space, drawn from a time-dependent probability distribution that changes both within and across days (see Supplementary Methods), and verified that repertoire dating can successfully recover the underlying structure of the models. The main parameters determining the relative alignment of the DiSC with the directions of within-day and across-day change are: k_a, the amount of within-day change along the DiSC; g_n, the amount of within-day change orthogonal to the DiSC; and b_n, the amount of across-day change orthogonal to the DiSC. These parameters are expressed relative to the amount of across-day change along the DiSC (thick black arrow in **a**, **b**). The two models shown (**a**, **c**) are characterized by different amounts of overnight consolidation of within-day changes along the DiSC. **a**, Model 1. Within-day change is aligned with the DiSC ($g_n = 0$) and is large ($k_a = 5$). The component of across-day change orthogonal to the DiSC is as large as the component of across-day change along it. In this scenario, overnight consolidation of within-day changes along the DiSC is weak for typical renditions (20% of change is consolidated, corresponding to a consolidation index of −0.8 in Fig. 1c, d). **b**, Repertoire dating percentiles for model 1, analogous to Fig. 3b. The time course of the 50th repertoire-dating percentile (typical renditions, red) closely reproduces the dynamics of change along the DiSC implied by **a**: within-day change along the DiSC is large (the red line extends over about five days) and consolidation is weak (the starting point on day $k + 1$ relative to day k moves by about 20% of the overall within-day range). **c**, Model 2. Within-day change has a large component orthogonal to the DiSC, whereas across-day change is aligned with the DiSC. In this scenario, overnight consolidation of within-day changes along the DiSC is strong (80% of the change is consolidated; consolidation index −0.2) for typical renditions. **d**, Repertoire dating percentiles for model 2, analogous to **b**. The time course of the 50th repertoire dating percentile (typical renditions, red) closely reproduces the dynamics of change along the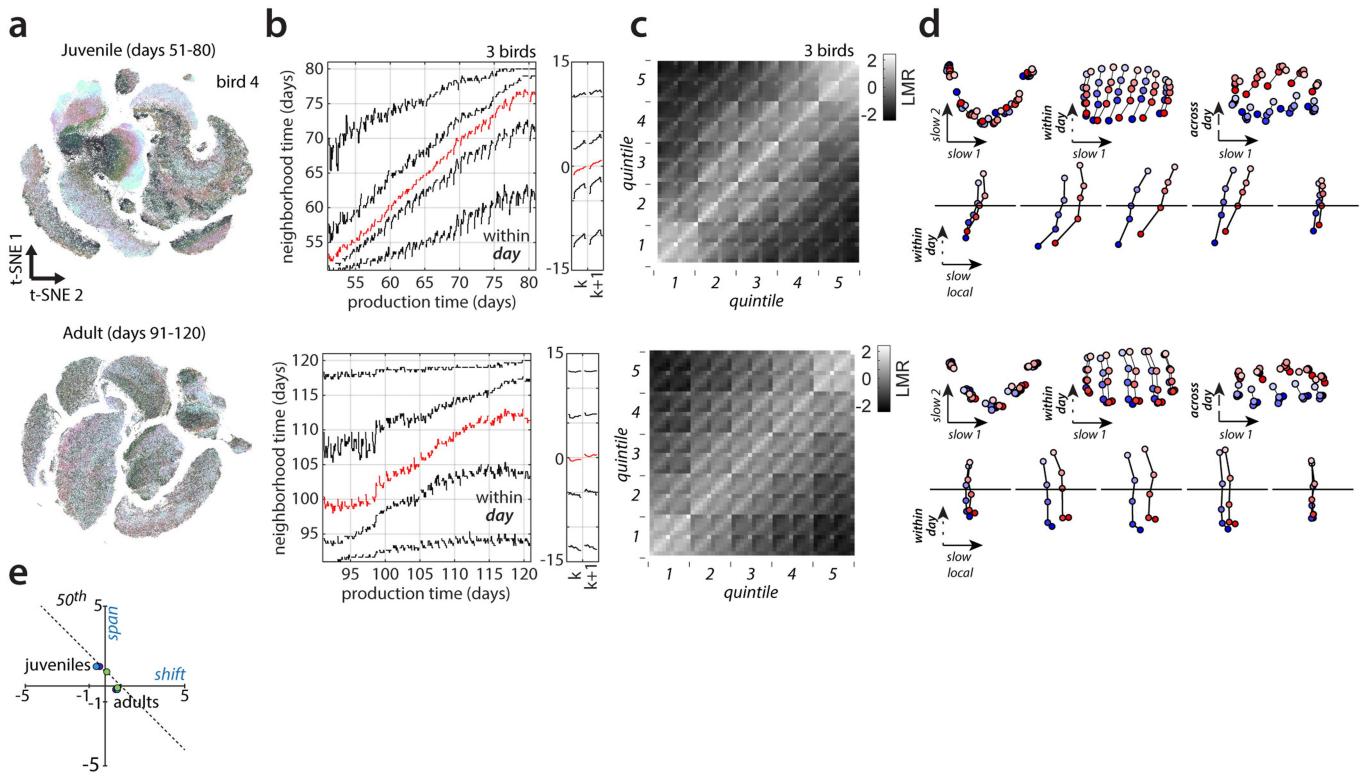 DiSC implied by **c**. In **b**, **d**, differences between anticipations (95th percentile) and regressions (5th percentile) correctly reflect the underlying model parameters (see Supplementary Methods). **e–j**, Validation of stratified behavioural trajectories. We generated three sets of stratified behavioural trajectories that differ with respect to the alignment of within-day and across-day change with the DiSC. We built each set of trajectories by arranging 50 points (five strata per day, five production time periods per day, on two consecutive days; same conventions as Fig. 3f, g) within a four-dimensional space. We then generated simulated stratified mixing matrices (**e–g**, replotted from Fig. 3f) by computing pairwise distances between all points, and transforming distances into similarities. We visualize the behavioural trajectories (**h–j**) with the same two-dimensional projections as in Fig. 3h–j, with the same scale along all dimensions. In all models, overnight consolidation along the DiSC is perfect (strong consolidation) for all strata. **e**, Model 1: within-day change and across-day change occur only along the DiSC. For each stratum (that is, each of the five 10-by-10 squares along the diagonal), similarity decreases smoothly with time, reflecting the gradual progression of the trajectory along the DiSC within and across days. **f**, Model 2: within-day change has a large component that is not aligned with the DiSC. **g**, Model 3: both within-day and across-day change have large components that are not aligned with the DiSC. The misaligned component of across-day change reduces the similarity between day k and day $k + 1$ compared with model 2, resulting in smaller values in the 5-by-5 squares comparing points from day k and day $k + 1$. **h**, Behavioural trajectories for model 1: the two-dimensional projection containing the DiSC (top) explains all the variance in the trajectories. **i**, Behavioural trajectories for model 2: similar to **h**, but points from different periods during the day are also displaced along an orthogonal direction of within-day change (middle). **j**, Behavioural trajectories for model 3: similar to **i**, but points from adjacent days are also displaced along an orthogonal direction of across-day change (bottom). Note that the models in **e–j** are implemented differently to the models in **a–d** (see Supplementary Methods).

Extended Data Fig. 5 | Repertoire dating control analyses. a–c, Within-bout effects, analogous to Fig. 3c, d. **a,** Within-bout effects computed only from renditions that fall into short bouts (the bout length is less than the median). **b,** Analogous to **a,** but computed only from renditions that fall into long bouts (the bout length is more than the median). The changes in the behavioural repertoire observed within a bout are qualitatively similar for short and long bouts (compare **a, b;** within-bout effects are most pronounced after day 70). In particular, the song becomes more regressive shortly before the end of a bout (5th percentile, bottom curves). This suggests that the analogous effect in Fig. 3c, d occurs at the end of a bout, rather than at a fixed time after the beginning of a bout. **c,** Analogous to Fig. 3c, d but computed over the entire dataset without prior clustering into syllables. The changes in behavioural repertoire differ in several respects from those in Fig. 3c, d, which were computed on individual syllables and then averaged across syllables (see Supplementary Methods). Here, the increase in regressions at the bout end is less pronounced. Moreover, large within-bout changes also occur for anticipations early in development. Both differences may reflect changes in the relative frequency of renditions from each syllable (for example, introductory notes) sung throughout a bout. Such changes in frequency can affect the results in **c,** which were computed on the unclustered data, but not those in **a, b. d,** Within-day effects, analogous to Fig. 3a, b, but computed for individual syllables, and then averaged across syllables and animals. The changes in behavioural repertoire are qualitatively similar to those in Fig. 3a, b, which were computed using the unclustered data. This similarity implies that the dynamics along the direction of slow change in Fig. 3 cannot be explained by changes during the day in the relative frequency of renditions from each syllable. **e,** Analogous to Fig. 3a, b but computed after shuffling production times among all data points. Within-day changes of the percentile curves are small under this null hypothesis. The maximal span of within-day fluctuations is 0.2 days, compared with 3.71 for the unshuffled data in Fig. 3b. The total repertoire spread (5th to 95th percentiles) is around 40 days, compared with around 23 days for unshuffled data. The 50th percentile curve is flat, implying that the shuffled data do not undergo a systematic drift over time (that is, do not describe a DiSC). The vertical separation between percentiles, then, reflects the range of production times in the data, not the spread along the DiSC. The time course of the 5th and 95th repertoire dating percentiles should thus be interpreted as the progression of regressions and anticipations along the DiSC only over the range of repertoire times covered by typical renditions (that is, approximately the vertical range of the 50th repertoire dating percentile). **f,** Analogous to Fig. 3e but for different distance metrics (Euclidean; correlation; Euclidean after time warping) and feature representations (32 acoustic features; 1 acoustic feature (entropy variance)). See also Extended Data Fig. 9.
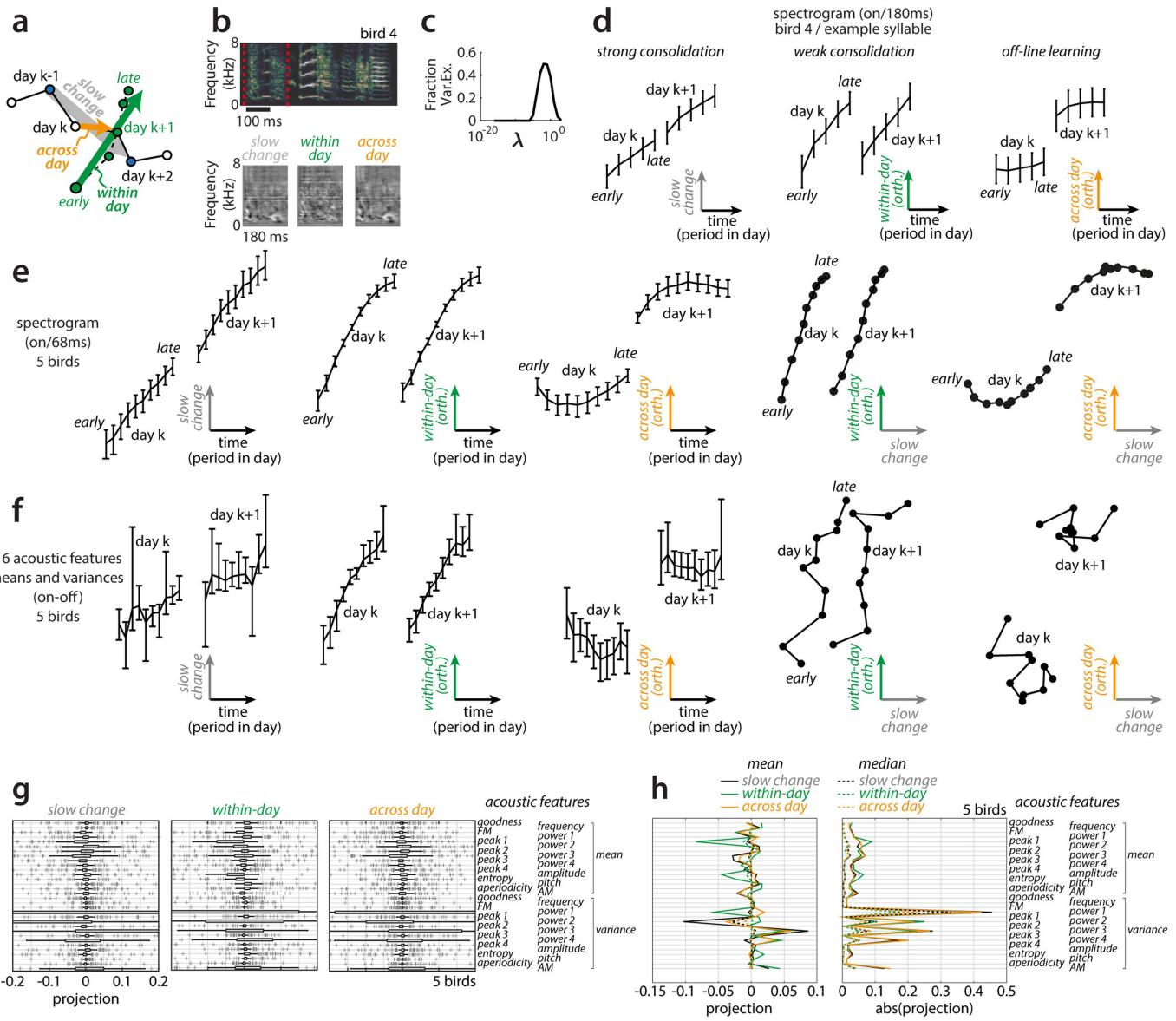
Extended Data Fig. 6 | Behavioural change in adult versus juvenile birds.
a–**d**, Comparison of within-day repertoire dating results during and after the
end of development (averaged over three birds). Top, juvenile birds; bottom,
same birds but as adults. **a**, Large-scale embeddings analogous to Fig. 2a. **b**,
Repertoire dating percentiles, analogous to Fig. 3a, b. **c**, Stratified mixing
matrix, analogous to Fig. 3g. **d**, Stratified behavioural trajectories, analogous
to Fig. 3h–k. **e**, Shift and span values for the 50th percentile, for juvenile and
adult birds. Points indicate individual birds. Song in adult birds is not static, but
the time course of change differs from that observed in juveniles. First, change
in adults is substantially less than in juveniles (see the slope of the 50th
percentile in the top versus bottom parts of **b**). Second, the relation of fast
(within-day) and slow (across-day) change differs in juveniles versus adults. In

juveniles, vocalizations move along the DiSC (y axes in **b**; slow local axis in **d**)
within each day and the repertoire time of typical renditions increases by about
one day from morning to evening (50th percentile; span is approximately one
day) and is maintained through the next morning (shift is approximately 0
days). In adults, typical renditions do not show within-day progress along the
DiSC (the span is approximately 0 days) but change overnight across days (the
shift is greater than 0 days). In adults, the regressive tail of the repertoire in
particular moves towards smaller values during the day (**b**, bottom right; 5th
percentile), whereas in juveniles it consistently moves towards larger values
(**b**, top). In both juvenile and adult birds, within-day change has a strong
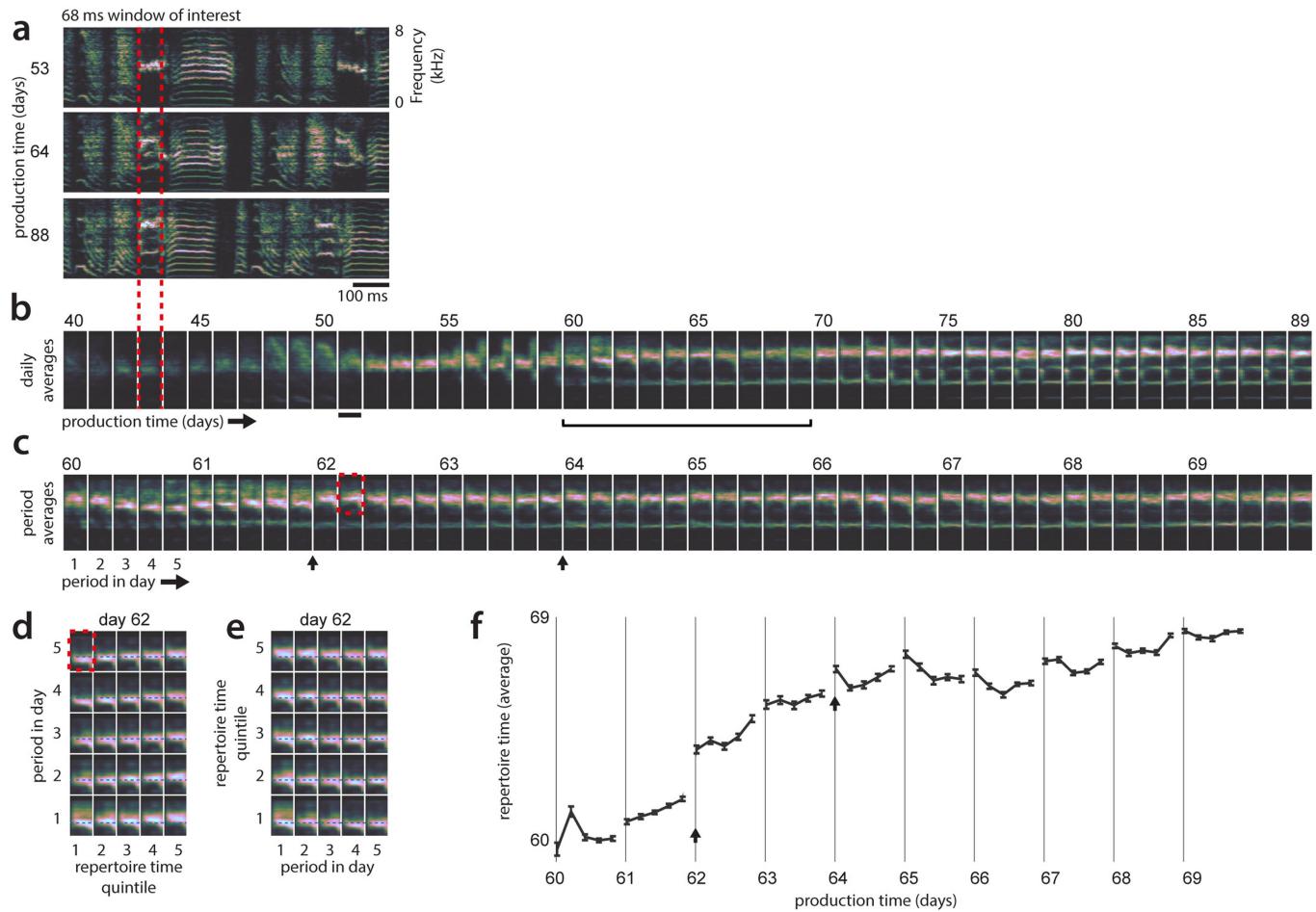component that is misaligned with the DiSC (within-day axis in **d**).

Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Local linear analysis. a–e, We validated the structure of change inferred with nearest-neighbour statistics (Fig. 3) with an approach based on linear regression in the high-dimensional spectrogram space (see Supplementary Methods). Unlike for the case of nearest neighbour-based statics, here each rendition must first be assigned to a cluster (that is, a syllable; compare with Fig. 2a) and each cluster is analysed separately. **a**, Illustration of the linearization scheme. First, we infer the (local) DiSC on days k and $k+1$ (grey arrow) as the vector of linear-regression coefficients relating production day to variability of renditions from days $k-1$ and $k+2$. Second, we infer the direction of within-day change (green arrow) as the linear-regression coefficients relating the period within a day to variability of renditions from days k and $k+1$, orthogonalized to the DiSC. Third, we infer the direction of across-day change (orange arrow) as the linear-regression coefficients relating production day to variability of renditions from days k and $k+1$, orthogonalized to the DiSC and within-day change. All three sets of coefficients, and the corresponding directions in spectrogram space, typically vary across days, syllables and birds. The progression of song along the DiSC and along the (orthogonalized) directions of within-day and across-day change are obtained by projecting renditions on day k and $k+1$ onto the corresponding directions. **b**, Example rendition of syllable b as in Fig. 1 (top, encapsulated by red lines) and inferred coefficients (directions in spectrogram space; bottom) for day $k = 57$. Bright and dark shades of grey mark spectrogram bins for which power increases or decreases, respectively, over the corresponding timescales in **a**. **c**, Dependency of cross-validated regression quality (fraction of variance explained; y axis) on the regularization constant (λ) for the estimation of the DiSC. One regularization constant was chosen for each syllable and the direction based on maximizing the leave-one-out cross-validation error on the training set. **d**, Progression of syllable b along the directions of change shown in **b**, during days 57 and 58. Renditions from each day are binned into ten consecutive periods on the basis of production time within the day (analogous to the ten periods in Fig. 3a, b; curves and error bars represent means and 95% bootstrapped confidence intervals). For simplicity of visualization, the time elapsed (x axis) during the night between days k and $k+1$ is not shown to scale. The position along the DiSC for the mo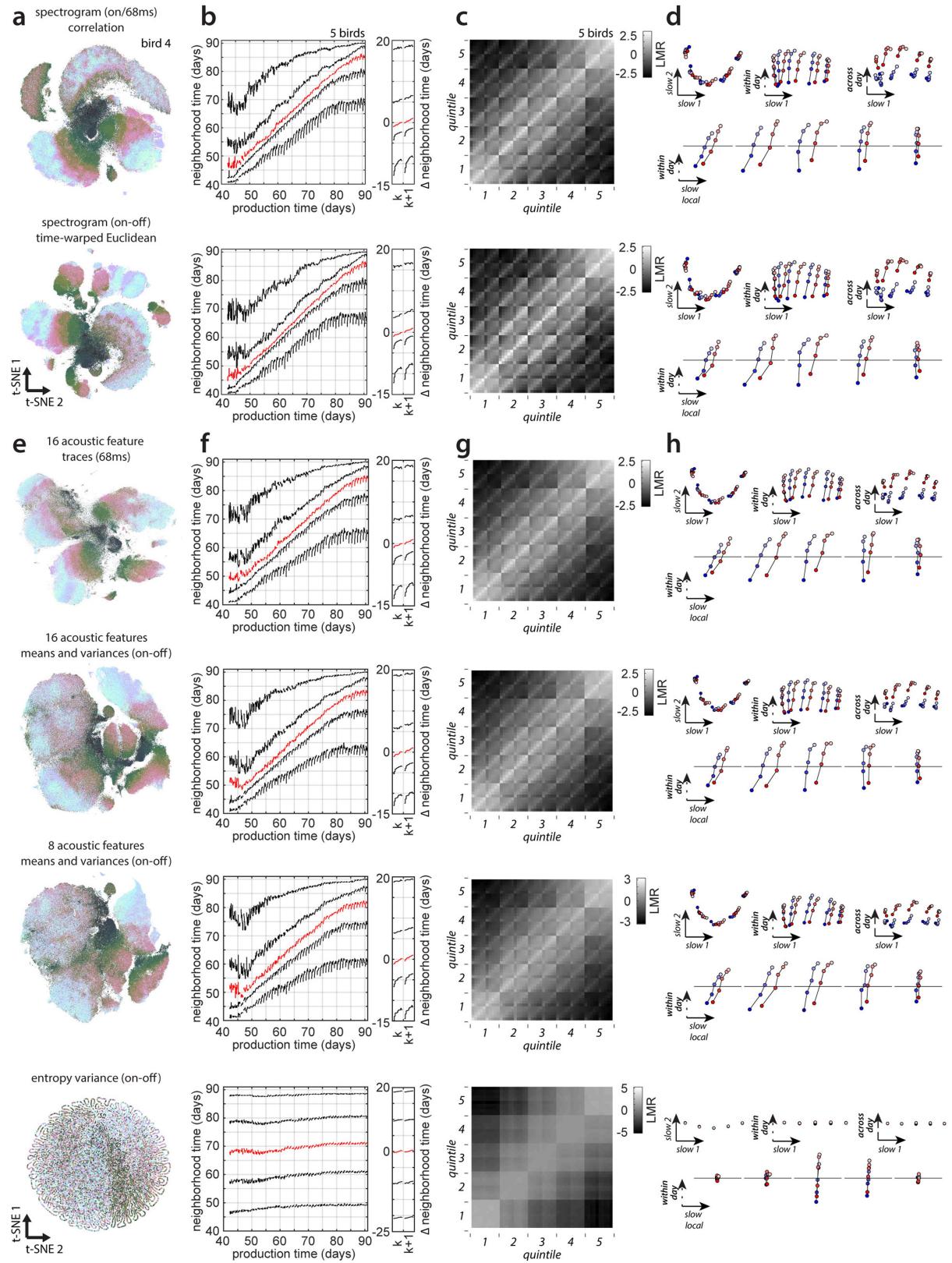rning of day $k+1$ is close to that for the evening of day k, indicating overall strong consolidation (left). The position along the direction of within-day change is reset overnight, implying that the underlying changes are not consolidated (middle). The position along the direction of across-day change jumps overnight, consistent with offline learning (right). We note that strong consolidation, weak consolidation and offline learning have all been reported previously, albeit in different behaviours and species[2,4,14,15,23,24]. The charts in **d** show that these different patterns of change can occur in the very same syllable along distinct spectral features (see also Fig. 1h and Extended Data Fig. 8). By considering features with different projections onto these directions, a wide range of consolidation patterns can be uncovered (see also Fig. 1h). **e**, As for **d**, but averaged across all four-day windows during days 60–69 and over all syllables and birds (same five birds as in Figs. 2, 3). The resulting averages include contributions from the entire behavioural repertoire, including regressions, typical renditions and anticipations. The two right-most panels show concurrent progression along the DiSC and the direction of within-day or across-day change, combining data from the first and second, or first and third, panels in **e**. These representations are analogous, and in qualitative agreement, with the behavioural trajectories in Fig. 3h–k (typical). **f**, Analogous to **e**, but computed on vocalizations represented by 32 acoustic features instead of spectrograms. Directions as in **e** can be retrieved, but progression along the DiSC appears noisier, suggesting that the 32 acoustic features do not fully capture in particular the slow spectral changes occurring over development (see also Extended Data Fig. 9). **g**, **h**, Contribution of individual acoustic features to the directions of slow, within-day and across-day change. As in **f**, the directions are computed in the space of 32 acoustic features. **g**, Distribution of coefficients in the retrieved orthonormalized directions. Thick and thin black bars represent means and 95% confidence intervals; crosses show outliers; thin vertical lines represent medians. **h**, Means (solid lines) and medians (dotted lines) of the signed (left) or unsigned (right) distributions in **g**. Most coefficients are small and variable, indicating that the alignment between any of the 32 acoustic features and the inferred directions of change is weak and highly variable over time, syllables and birds.

Article

Extended Data Fig. 8 | Behavioural variability and stratification in an example syllable. a, Songs of an example bird for three days during development. Only those spectrogram segments that belong to a particular syllable and location in the motif (68-ms window of interest; red dotted lines) are analysed in the subsequent panels. **b**, Developmental changes over the course of weeks. Renditions are binned by production day, and averaged. The most apparent changes are an increase in pitch and the later successive appearance of additional spectral lines at low frequencies. **c**, Within-day and across-day changes for days 60–69. Renditions are binned into five production-time periods spanning a day and averaged within bins. On many days, the changes within a day do not appear to recapitulate the changes occurring across days (for example, days 60 and 65; within-day progression does not smoothly transition between the vocalizations on preceding and subsequent days; see **b**). The averages also reveal occasional overnight 'jumps' in the properties of the vocalizations (see, for example, the vertical black arrows).

d, Comparison of within-day change and change on longer timescales. Renditions within each period and day were split into strata according to their repertoire times (for example, the quintiles in Fig. 2d), resulting in 25 averages, one for each combination of stratum and period within the day. Only the upper part of the spectrogram is shown (red rectangle in **c**). The progression along strata (x axis) emphasizes the large extent of motor variability along the DiSC existing within a single day (day 62). **e**, Same averages as in **d**, but with x and y axes swapped. In particular for regressive renditions (quintile 1), change within day 62 (x axis) does not recapitulate developmental changes occurring over months (x axis in **d**). **f**, Repertoire dating based on repertoire time (as in Extended Data Fig. 3c). Each point corresponds to a production-time period and the average of all repertoire times of renditions in that period. Error bars show bootstrapped 95% confidence intervals. The change in repertoire time, which is computed without using a low-dimensional parametrization of vocalizations captures the movement along the DiSC.

a spectrogram (on/68ms) correlation

bird 4

spectrogram (on-off) time-warped Euclidean

t-SNE 1
t-SNE 2

b 5 birds

neighborhood time (days)
production time (days)

Δ neighborhood time (days)
k k+1

c 5 birds

quintile
quintile

LMR
2.5
0
−2.5

d

slow 2 / slow 1
within day / slow 1
across day / slow 1

within day / slow local

e 16 acoustic feature traces (68ms)

16 acoustic features means and variances (on-off)

8 acoustic features means and variances (on-off)

entropy variance (on-off)

t-SNE 1
t-SNE 2

f

neighborhood time (days)
production time (days)

Δ neighborhood time (days)
k k+1

g

quintile
quintile

LMR
2.5
0
−2.5

3
−3

5
0
−5

h

slow 2 / slow 1
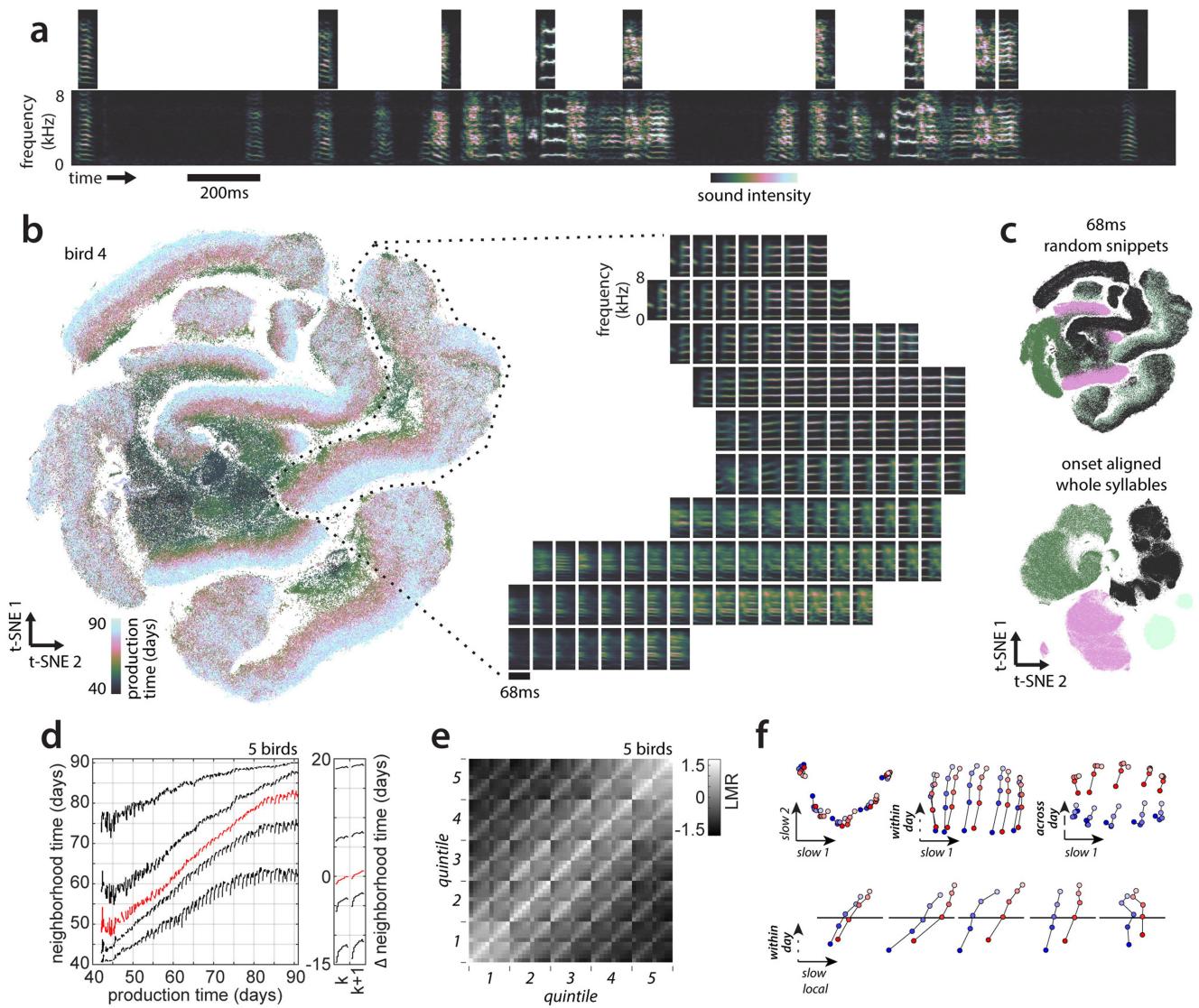within day / slow 1
across day / slow 1

within day / slow local

Extended Data Fig. 9 | See next page for caption.

Article

Extended Data Fig. 9 | Behavioural change based on alternative distance metrics and features. To demonstrate the robustness of the proposed nearest-neighbour statistics, we verified that the inferred time course of behavioural change is reproduceable using a number of different distance metrics (used to define nearest neighbours) and parameterizations of vocalizations. **a**–**d**, We recomputed the main analyses using a Pearson's correlation metric on 68-ms onset-aligned spectrogram segments (first row); and the Euclidean distance on onset-to-offset spectrogram segments that were linearly time-warped to a duration of 100 ms (second row). For comparison, the main analyses in the text were based on Euclidean distance on 68-ms onset-aligned spectrogram segments (for example, Fig. 2c–f, 3). **a**, *t*-SNE visualization based on the corresponding distance metrics and sound representation for the example bird, analogous to Fig. 2a. **b**, Repertoire dating averaged over birds, analogous to Fig. 3a, b. **c**, Stratified mixing matrices averaged over birds, analogous to Fig. 3g. The mixing values are highly correlated across distance metrics: Euclidean (main text) versus correlation, variance explained = 92%; Euclidean (main text) versus time-warped Euclidean, 93%. **d**, Stratified behavioural trajectories based on **c**, analogous to Fig. 3h–k. The results in **a**–**d** are consistent with those in Fig. 3, showing that our findings are robust with respect to the exact definition of nearest neighbours. Moreover, the overall structure of the behavioural trajectory appears to depend only minimally on changes in tempo and spectrogram magnitude (first row: Pearson's correlation is invariant to changes in overall magnitude of vocalizations; second row: time-warped Euclidean distance is invariant to changes in tempo). **e**–**h**, We recomputed all main analyses with four additional parameterizations of vocalizations: time-dependent normalized acoustic feature traces for 16 acoustic features within 68-ms windows after syllable onset (first row); means and variances of the same 16 acoustic features over entire syllables (second row); means and variances of 8 of the 16 acoustic features (third row); and a one-dimensional parametrization consisting solely of entropy variance computed over entire syllables (fourth row). Feature means and variances were *z*-scored across all syllables. For all of these parameterizations we defined nearest neighbours with the Euclidean distance. **e**, Embedding using *t*-SNE based on the corresponding parameterization and metric. For entropy variance alone, the embedding appears locally one dimensional (for visibility, data points are larger than for the other parameterizations). Entropy variance maps mostly smoothly onto this one-dimensional manifold (data not shown). **f**, Repertoire dating averaged over birds, analogous to Fig. 3a, b. Repertoire dating based on entropy variance alone fails to reproduce most of the results in Fig. 3 obtained with spectrogram segments. The percentile curves are almost flat, indicating that renditions cannot be reliably assigned to their production times on the basis of entropy variance alone. In this case, vertical separation between percentiles cannot be interpreted as spread along the DiSC (see Extended Data Fig. 5e). For entropy variance alone, span is greater than zero across all percentiles, but consolidation is consistently close to zero. **g**, Stratified mixing matrix averaged over birds, analogous to Fig. 3g. The match with the mixing matrix in Fig. 3g decreases as the dimensionality of the parameterization is reduced (spectrogram versus time-dependent feature traces: variance explained = 93%; spectrogram versus 16 acoustic feature means and variances, 91%; spectrogram versus 8 acoustic feature means and variances, 84%; spectrogram versus entropy variance, 54%). **h**, Stratified behavioural trajectories based on **g**, as in Fig. 3h–k. The inferred behavioural trajectories are similar across the first three song parameterizations. However, these alternative parameterizations result in more vertical separation between percentiles in **f**, suggesting that they capture the direction of slow change less well (compare with Fig. 3a and Extended Data Fig. 5e). Parameterizations of reduced dimensionality also result in progressively less defined syllable clusters in the embeddings (**e**, top to bottom). These observations suggest that a parameterization based on the full spectrogram is better suited to capture the different directions of change explored during development (see also Extended Data Fig. 7). Note that for entropy variance (bottom row), the projections onto the local direction of slow change are highly magnified compared with the projections in the top panels.

Extended Data Fig. 10 | Behavioural change based on random spectrogram segments. We recomputed all main-text analyses with a random segmentation of behaviour that does not require alignment to syllable onsets. This segmentation scheme can be applied to behaviour that does not fall into temporally discrete elements. Here each data point corresponds to a randomly chosen 68-ms spectrogram snippet drawn from a period of singing. Not all song was sampled, as we used 1,000,000 non-overlapping segments for each bird (see Supplementary Methods). **a**, Vocalizations of the example bird (as in Fig. 1a) from day 76, with example segments used for the analysis (at the top). **b**, *t*-SNE visualization for random segments from the example bird, based on nearest neighbours defined with respect to the Euclidean distance (left) and average spectrograms for different locations in the (*t*-SNE) embedding (right; analogous to Fig. 2b). Clusters corresponding to individual syllables are elongated compared with Fig. 2a. Variation along one direction within the cluster tends to account for production time (colour bar), while variation along another direction tends to reflect the timing of segments relative to syllable onsets. **c**, Embedding from Fig. 2a (bottom) and embedding of random 68-ms segments (top). Points in both embeddings are coloured according to cluster identity defined on onset-aligned spectrogram segments covering entire syllables (bottom). The colour of each point corresponding to a random snippet (top) corresponds to the cluster identity of the surrounding syllable. Some clusters in the embedding based on random segments contain points assigned to two different syllables (for example, black versus green colours). **d**, Repertoire dating averaged over birds, analogous to Fig. 3a, b. **e**, Stratified mixing matrix averaged over birds, analogous to Fig. 3g. The mixing values are highly correlated with those in Fig. 3g (variance explained = 89%). **f**, Stratified behavioural trajectories based on **e**, as in Fig. 3h–k. The results in **d**–**f** largely reproduce the corresponding findings obtained with onset-aligned 68-ms spectrogram segments (Fig. 3) as well as with other song parameterizations (Extended Data Fig. 9). Nonetheless, the overall effect sizes are reduced, probably because of the additional variability introduced by the random position of segments relative to syllable onsets. In **d**, the vertical separation between the 5th and 95th percentiles is increased and the slope of 50th percentile is reduced compared with the main-text analyses (Fig. 3a), suggesting a noisier representation of the direction of slow change (see Extended Data Fig. 5e) compared with onset-aligned 68-ms segments (Fig. 3).

nature research

Corresponding author(s): *VALERIO HANTE* (handwritten)

Last updated by author(s): YYYY-MM-DD *2019_09_10* (handwritten)

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection   Labview, Matlab

Data analysis   Matlab

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For each animal all available data in the given developmental window of interest was chosen.
Data exclusions	Out of a set of 16 animals, 11 were excluded because recording conditions were inadvertently changed mid-experiment by modifying the amplifier settings.
Replication	We did not attempt replication using a different dataset.
Randomization	n/a
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Zebra Finches, Male
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	All experimental procedures were approved by the Veterinary Office of the Canton of Zurich.

Note that full information on the approval of the study protocol must also be provided in the manuscript.