# Cultural conformity and persistence in the context of differing site fidelity

Timothy H. Parker[1]
Bridget Sousa[2,3]
Stephan T. Leu[4,5]
Stacy Edmondson[6]
Cecily Foo[1,7]
Amy Strauss[1,8]
Hanna Kahl[1,9]
Kristen Ballinger[1,10]
Eric Ross[11]
Mareile Große Ruse[12]
Maria Sandsten[13]
Bram H. F. Verheijen[14]
William Jensen[15]

[1]Department of Biology, Whitman College, Walla Walla, WA 99362, USA
[2]Department of Biology, University of Kentucky, Lexington, KY, 40506, USA
[3]Swaim Biological Inc., 4435 First Street, PMB 312 Livermore CA 94551, USA
[4]Department of Biological Sciences, Macquarie University, Sydney, New South Wales, 2109, Australia
[5]School of Animal and Veterinary Science, University of Adelaide, South Australia, 5005, Australia
[6]Department of Mathematics and Statistics, Whitman College, Walla Walla, WA 99362, USA
[7]West, Inc. Golden Valley, MN, 55427, USA
[8]Graduate Program in Organismic & Evolutionary Biology, University of Massachusetts, Amherst, MA 01003, USA
[9]Department of Entomology, University of California – Davis, Davis, CA, 95616, USA
[10]4215 Dayton Ave N, Seattle, WA, 98103, USA
[11]Department of Applied Environmental Science, California State University - Monterey Bay, Seaside, CA, 93955, USA
[12]Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark
[13]Mathematical Statistics, Centre for Mathematical Sciences, Lund University, Lund, Sweden
[14]Kansas Cooperative Fish and Wildlife Research Unit, Kansas State University, Manhattan, KS 66506, USA
[15]Department of Biological Sciences, Emporia State University, Emporia, KS, USA

## Abstract

Animal culture often shows geographic structure, with nearby individuals sharing more cultural features than individuals further apart. However, spatial extent of cultural features, along with the degree of conformity to local cultures, vary within and among species. Further, rates of cultural change presumably also vary, though documentation of temporal variability lags behind documentation of spatial variability, and mechanisms driving this variation have not been sufficiently explored. We hypothesized that conformity to local culture in Oscine songbirds and the persistence of culture over time and space are promoted by habitats that facilitate stable populations in which individuals show relatively high site fidelity. In contrast, sites in which habitat features cause rapid population turnover provide more vacant territories and so more opportunities for colonization. Colonization should drive more rapid cultural change, either through adult colonists importing foreign cultural variants or young colonists making errors as they learn the local song. To test these hypotheses, we examined temporal and spatial variation in vocal culture in a songbird (dickcissel, *Spiza americana*) in two distinct habitats. As predicted, we found high average site fidelity in relatively stable native grasslands and much lower average site fidelity in nearby cropland sites which were disturbed by farming practices during the breeding season. We also found higher levels of average song similarity and slower average changes in vocal culture in our grasslands relative to croplands, though we found no evidence of different spatial scales of song similarity between these habitats. Although our conclusions are based on many recorded songs, automated assessments of song similarity underestimated the overall degree of song similarity. Thus we may have also underestimated the strength of the effects of time and distance on song similarity. Despite these shortcomings, our results call attention to the understudied but potentially important role of demographic factors influencing cultural evolution.

keywords:

apparent survival, dialect, dickcissel, site fidelity, song sharing

## Introduction

Among the most widespread and conspicuous forms of (non-human) animal culture are the geographically variable songs of many Oscine songbird species (Lemon 1975, Krebs and Kroodsma 1980, Baker and Cunningham 1985, Aplin 2019). Understanding the substantial variation in conformity, geographic extent, and temporal stability of these vocal cultures is an interesting challenge for behavioral ecology. Researchers have been making progress describing and explaining this variation for decades (e.g., McGregor and Thompson 1988, Payne et al. 1988, Kroodsma et al. 1999, Beecher and Campbell 2005, Derryberry 2009, Ju et al. 2019), especially in describing differences among species (Beecher and Brenowitz 2005, Handley and Nelson 2005), but important potential drivers of cultural variation remain poorly explored. In particular, differences among populations within species in cultural conformity, geographic extent of culture, and temporal stability of culture, are not well understood. We know these differences exist among populations within species (e.g., Mundinger 1975, Bitterbaum and Baptista 1979, Kroodsma and James 1994, Tracy and Baker 1999, Lachlan and Slater 2003, Schook et al. 2008, Ju et al. 2019), but the most well-understood system are those less common ones in which culturally divergent populations are geographically isolated (e.g., Laiolo and Tella 2005, 2007). The more common condition of a cultural mosaic in continuous populations has been largely unexplained (but see, Ortega et al. 2014).

This variability in spatial or temporal patterns of song culture likely has multiple causes, but we are particularly interested in the possibility that variation in cultural patterns may be influenced by variation in population processes, such as the movement of individuals (Fayet et al. 2014, Aplin 2019), which might differ across landscapes. For instance, territory vacancies, created by emigration or by death, should influence song culture (Baker and Thompson 1985, Kroodsma et al. 1985, Goodfellow and Slater 1986, Holland et al. 1996). In the most extreme cases where all individual members of a given culture either leave or die and a site is colonized by a new set of individuals, we would expect dramatic and immediate changes in song type and geographic patterns of song culture (Holland et al. 1996). This extreme scenario may be rare, but variation in rates of population turnover within a species is common, presumably because of variation in features of the habitat (Winker et al. 1995, Newton 1998), and might have major implications for vocal traditions. The manner in which population turnover influences vocal culture probably depends on features of the vocal culture. In the only study we know of to explicitly address the relationship between population turnover and vocal culture, the study species was one in which individual birds sing multiple different song types learned from multiple tutors and local vocal neighborhoods are characterized by the presence of diverse song types (Ortega et al. 2014). In this system, when poor habitat quality reduced site fidelity, vocal diversity declined because the new arrivals had fewer tutors, and thus fewer song types, to imitate. However, in systems characterized by vocal conformity rather than diversity, we might expect the opposite outcome (Ortega et al. 2014). For instance, in systems normally characterized by conformity, the larger the proportion of territories occupied by new birds (young individuals or immigrants attempting to match the local song) in a given year, the higher the proportion of individuals learning the local song that year. Making the typical assumption of a constant probability of copying errors per bird learning new song (Williams and Slater 1990, Lachlan and Slater 2003), the more individuals that are learning song each year, the greater the number of errors that accumulate in the population (Goodfellow and Slater 1986). These deviations from tutors' songs, whether due to improvisation or learning error (Beecher and Brenowitz 2005), are presumably the ultimate source of new song types (Slater 1986). Thus, a larger number of novel song variants should occur in areas with low site fidelity, and this could lower the local level of song similarity (cultural conformity). In some species, territories made available by death or emigration might also be filled by immigrant adults that bring song learned elsewhere, which could also lower local similarity and increase the rate of cultural change (Goodfellow and Slater 1986, Williams and Slater 1990). Further, not only might new song variants emerge or be introduced more rapidly in populations with high turnover, temporal change in song types within these neighborhoods could also be accelerated since elimination of song types due to drift should be more likely with few individuals persisting from one year to the next (Goodfellow and Slater 1986, Nunn et al. 2009). Therefore, if we wish to understand variation in cultural patterns in continuously occupied landscapes, we should study how habitat differences influence variation in territory turnover.

An excellent species for studying the role of population processes in influencing patterns of vocal culture is the dickcissel, a migrant songbird breeding in grassland and cropland landscapes of the Eastern Great Plains of North America (Temple 2002). The vast majority of dickcissel males each produce a single song (Figure S1), at least while occupying the same breeding territory within seasons (Schook et al. 2008; Author 1 unpublished data). Dickcissels show striking geographic structure in their song culture, with most males in close proximity to each other singing almost identical song phrases, and the degree of similarity of phrase types often declining considerably as the distance between recording sites approaches 2 km (Schook et al. 2008). Circumstantial evidence suggests that male dickcissels learn their song from immediate adult neighbors after dispersal, but that adult males occasionally move to different cultural areas and sing the song their learned on their prior territory (see supplement). Thus we expected high territorial turnover in dickcissels to increase the number of new song learners and immigrants with new song types and thus accelerate cultural change. Further, we had a strong expectation that sites covered primarily by crops would have lower site fidelity than sites covered primarily by native grasslands. This is because dickcissel males tend to abandon territories in response to the loss of dense vegetation cover and due to reproductive failure (Schartz and Zimmerman 1971, Zimmerman 1982, Zimmerman and Finck 1989, Frawley and Best 1991, Igl 1991), both of which result from crop harvest during the dickcissel breeding season.

We tested a set of hypotheses associated with spatial and temporal change in the vocal culture of dickcissels (see supplement for a description of the evolution of this project and the timing of the development of our hypotheses). In summary, we hypothesized that accumulation of imperfect song copies and/or the importation of song variants from elsewhere change local song culture over time and space in dickcissels. Further we hypothesized that these changes are more rapid in crop landscapes than in grassland landscapes because population turnover is faster in cropland due to crop harvest during nesting. We derived several testable predictions from these hypotheses:

- average pairwise song similarity within a site will decline as the number of years between recordings increases
- average pairwise song similarity within a site will decline as the geographic distance between the recorded individuals increases
- adult male dickcissels in croplands will show lower apparent survival (which is a combined function of site fidelity and true survival), than those in grassland
- overall levels of song similarity in dickcissels will be higher in the habitat with the higher apparent survival
- habitat with lower apparent survival of male dickcissels will show steeper declines in average pairwise song similarity as a function of the number of years between recordings relative to the habitat with higher apparent survival

- habitat with lower apparent survival of male dickcissels will show steeper declines in average pairwise song similarity as a function of geographic distance between recorded individuals relative to the habitat with higher apparent survival

## Methods

### Field methods

We recorded songs and location data from both marked and unmarked dickcissels at seven sites in northeastern Kansas between 2006 and 2014 (Tables S1 – S3). We did not visit all sites in all years. In native perennial grassland we had four non-adjacent sites on the Konza Prairie Biological Station (KPBS). We also worked at three cropland sites with a mix of crops (primarily winter wheat and alfalfa) and weedy field edges within 40 km of KPBS. In some years we monitored sites throughout the breeding season (from late May or early June into late July), but in some other years, we made only a single site visit during the peak of breeding between late May and mid-June). Field observers were not blinded to the hypotheses, but relevant measurements were not made in the field, so the lack of blinding is unlikely to have introduced bias. Years with only one visit could not contribute to our estimate of within-season apparent survival or to within-male comparisons of songs across dates within years.

In years in which we monitored a given site throughout the breeding season, we attempted to capture and mark all territorial male dickcissels present at that site, though typically a subset eluded capture. To capture a targeted individual, we played dickcissel song in the vicinity of mist nets placed near territorial song perches. Immediately upon capture we fitted male dickcissels with a uniquely numbered aluminum leg band and three (occasionally fewer in 2006) colored leg bands in a unique combination for later identification without recapture. By color banding birds and then observing these banded individuals from afar we were able to determine the identity, presence, and distribution of individual birds during and between years.

We attempted to record the song of all male dickcissels, banded and unbanded, on all sites monitored in a given year. We recorded most songs onto digital media using Marantz PMD680 recorders and Sennheiser ME66 shotgun microphones [Author 1 field teams], though we recoded some onto analog tapes using a Sony TCM-5000EV cassette recorder and Sennheiser ME67 shotgun microphone, later digitized using Raven 1.3 with default settings [Author 2 field teams]). For each male on each visit to a site, we attempted to record at least one clear song with limited background noise. Because an individual male sings only a single version of his song and a complete song lasts between one and two seconds (Schook et al. 2008), when our target bird was close and ambient noise was low we sometimes recorded for as briefly as 20 seconds to obtain several clearly recorded songs and our narration. However, various forms of noise, including the singing of other birds, usually led us to record for periods of one to five minutes or more to be confident that we recorded the single clear song that we needed for our analysis. The observers making the recording typically vocally identified each song by the target male on the recording immediately after each song was sung so that it could be distinguished from any other nearby conspecific songs. For banded birds, we generally confirmed the band combination both before and after recording. For unbanded birds, we only attempted to record all individuals at a site on one day of the season so that we could be confident that we were not recording any individuals twice. For all recorded birds, banded or unbanded, we recorded location coordinates with a GPS within 10 m of the location where the focal individual had perched while singing. For banded birds, we averaged multiple GPS locations across dates to generate an average location for a given year.

## Mark-recapture analyses

To assess whether male dickcissels show higher site fidelity in the perennial grassland sites relative to the cropland sites, we used data from re-sighting of our banded males within and across seasons (Table S2). We analyzed encounter histories with Cormack-Jolly-Seber models for live encounter data available in MARK (version 8.1; White and Burnham 1999), and tested for differences between cropland habitat and grassland habitats in both apparent survival (ϕ) and detection probability (*p*). We expect that apparent survival over two-week periods (the typical duration between our site visits) within breeding seasons is driven primarily by emigration, and so consider within-season apparent survival to represent site fidelity. Apparent survival over the non-breeding season can also be influenced by site fidelity if individuals survive but choose a different breeding location in the subsequent year. However, regardless of the extent to which differences in apparent survival between sites are driven by within-season or between-season emigration, or the extent to which they are driven by differences in emigration as opposed to differences in true survival, reduced apparent survival drives an increase in turnover. With greater turnover, there are more opportunities for cultural evolution with the copying of local song or the introduction of foreign song by immigrants.

We derived our mark-recapture data from two separate studies, one conducted by the Author 1 team, and the other by the Author 2 team. Some individual birds were monitored by both teams. In these cases, we included only the data from the Author 1 team so that we could test for differences in estimates derived by the two different teams (see below) without double counting birds, and so that we could maximize the sample size of the Author 1 data set, which included birds from both grassland and cropland sites. Because these two studies differed in capture and re-sighting methods, we tested for differences in their estimates of ϕ and *p* while simultaneously comparing estimates of ϕ and *p* between cropland and grassland habitats. The Author 1 team monitored banded male dickcissels at both grassland (n = 102 marked individuals, of which 81 were resighted at least once) and cropland (n = 46 individuals, 34 of which were resighted at least once) sites. In years when this team banded, they captured and marked birds for about two weeks beginning the last week of May, which is typically at least two weeks after male dickcissels begin establishing territories. These researchers then attempted to re-sight banded individuals approximately once every two weeks until late July. The Author 2 team monitored banded male dickcissels only in grasslands (n = 109 individuals, excluding individuals also monitored by Author 1 team, of which 81 were resighted at least once), captured birds for banding throughout the season, beginning in early May, and attempted to re-sight banded birds most days until early August. To assess the differences between the two studies, we compared models that divided capture histories among three categories (1. cropland, 2. Author 1 grassland, 3. Author 2 grassland [3hab]) vs. models that pooled data from the two studies and thus allocated all capture histories to either 1. cropland or 2. grassland (2hab). We also assessed whether apparent survival differed between breeding and non-breeding seasons (ssn). Last, to obtain an accurate estimate of apparent survival within the breeding season, we tested for a difference in detection probability between the first and subsequent survey rounds within the Author 1 dataset, to account for the fact that an experienced observer (Author 1) always participated in the first round, but not in later rounds (rnd). We did not assess among-year variation in detection probability or apparent survival in our data because the mark-resighting data were too sparse. Our global model therefore contained a habitat and season effect on apparent survival (3hab x ssn) and a habitat and a survey round effect on the detection probability (3hab x rnd; Table1). We fitted all models with design matrices and a logit-link function. To avoid bias in parameter estimates, we generated weighted average estimates and associated standard errors across all models using Akaike weights, and thus accounted for both sampling and model-selection uncertainty (Burnham and Anderson 2002).

We were not able to test the global model for overdispersion via the c-hat procedure in MARK because of missing survey rounds in certain years. Instead, we assessed the potential impact of controlling for overdispersion by manually increasing the variance inflation factor ($\hat{c}$) and assessing the resulting model ranking (Table S5). Our top set of models ($\Delta$AICc < 2) all contained the same habitat variable (3hab) for both apparent survival and recapture probability for $\hat{c} = 1 – 4$, equivalent to overdisperion that ranges from absent to high (Burnham and Anderson 2002). This variable (3hab) was missing for some models in the top set ($\Delta$AICc < 2) for apparent survival only when we set $\hat{c}$ at or above 5, a level that would indicate major problems with model fit (Burnham and Anderson 2002). However, the variable 3hab was retained in all models of detection probability for all values of $\hat{c}$. Further, at all values of $\hat{c}$, the seasonal effect (ssn) on apparent survival was included. Thus, we conclude that our model selection results were robust to overdispersion.

## Song processing prior to analyses

For our primary analyses, we selected one song per male per year, but if the same banded male was present at a site in multiple years, we included one song from each year available. We chose to include the same males from multiple years because (a) an important component of cultural stability is likely the persistence of individuals across years, (b) individual males can change their song from one year to the next (unpublished data) and so retention of males across years need not necessarily increase estimates of song similarity across years, and thus (c) a valid description of the degree of song stability over time should include as many males present in each year as possible.

Prior to assessing similarity among recorded songs, we processed songs in Raven (version 1.4) (Bioacoustics_Research_Program 2011) to optimize comparability. We first identified a song in which all components were evident and high amplitude background noise did not overlap the song’s notes. If no single song met these criteria, we isolated a sufficient number of songs (typically no more than 4) such that we could obtain a clear and unobscured version of each separate song phrase from among the selected renditions. We then filtered out noise below 2000 and above 20000 kHz. Most background noise is low frequency, and so filtering below 2000 kHz dramatically reduced noise. However, in most recordings, notes produced by other individual birds, by insects, or by various other sources in the environment remained in the recording. We could readily distinguish the notes of our target bird from these background notes by observing longer stretches of the recording to see the pattern produced by the target bird and the sources of background noise. We filtered out any background noise that approached the amplitude of the target bird’s quietest notes by using the select option in Raven and filtering out the content of the selections.

## Creating a song similarity score

We created a song similarity score between all pairs of selected songs within each site, including across years. This score combined an assessment of the proportion of phrase types shared between individuals multiplied by a quantitative assessment of similarity between phrases of the same type. The index ranged from 0 to 1, with 1 indicating the two individuals sang exactly the same song types with exactly the same arrangement of notes. We explain below how we generated these two pieces of information and combined them into a single, intuitive score of song similarity.

First, we classified all phrases in our target recordings into categories or types and quantified the proportion of shared phrase types. The classification is straightforward because we can readily divide all dickcissel song phrases into one of three categories. All dickcissel songs include phrases from at least two of these categories, and some contain all three (supplement). Based on the presence or absence of these phrase categories, we calculated Jaccard's Index of similarity (JI = number of phrase types sung by both individuals/(number of phrase types sung by both individuals + number of phrase types unique to bird 1 + number of phrase types unique to phrase 2), in which 0 = no overlap in phrase types and 1 = production of an identical set of phrase types. Because Jaccard's index represents the proportion of overlapping categories, we were able to multiply this value by the quantitative similarity averaged across shared phrase types (see below).

The quantitative similarity scores were based on comparisons between song phrases of the same type. We did not compare whole songs to each other because there is moderate variability within males and even greater variability among males within a local cultural area regarding the number of repetitions and timing of phrases, but high repeatability regarding the internal structure of phrase types. Thus focusing on individual phrases allows us to focus on the most consistent component of vocal culture, and it is this consistency we are seeking to quantify and explain.

We based the quantitative assessment of similarity within a phrase type on two distinct methods (implemented in MATLAB version R2016b; see https://(redacted for double blind review) for code). The first involved assessing the similarity of note positions as evident in sound spectrograms (Große Ruse et al. 2016, Sandsten et al. 2016). This multi-tapered spectrogram method compromises time and frequency resolution in order to reduce noise sensitivity. This is accomplished by using multiple windows or 'multitapers'. The second method, which is particularly useful in recognizing repeated patterns within phrases, is referred to as a filtered ambiguity spectrum (Große Ruse et al. 2016, Sandsten et al. 2016). This is derived with a Fourier transformation of the spectrogram in the two dimensions of time and frequency. Instead of marking the position or number of notes on the spectrogram, this method determines time and frequency differences between notes, and is thus robust to slight differences in the number or timing of repeated elements that are known to reduce estimates of similarity when focusing entirely on the spectrogram. Because these two methods provide different information regarding quantitative similarity, as part of the process of generating an overall average similarity score, we averaged the pairwise comparison values in the two similarity matrices (see supplement for more details, R versions 3.3.3 and 3.4.2, see https://(redacted for double blind review) for code). We then averaged these quantitative similarity scores across all shared phrase types for a given pair of songs.

Finally, we combined the Jaccard's Index of Similarity, which is the proportion of song types shared, with the quantitative measure of similarity between shared phrase types by multiplying the two to produce our overall song similarity score. This multiplication is justified because the quantitative score ranged from 0 to 1 with increasing similarity (as described in the supplement) for shared song phrases, but the score is discounted by the proportion of phrase types shared by multiplying by the Jaccard's Index, which also ranges from 0 to 1. Thus we generated an intuitive score of overall song similarity ranging from 0 to 1 between all pairs of individuals at each site (R versions 3.3.3 and 3.4.2, see https://(redacted for double blind review) for code).

Reliability of song similarity score

To evaluate the effectiveness of our method for assessing song similarity, we compared the levels of song similarity between two different recordings of the same individual bird in the same year (Table S4; within-individual) with recordings across individuals (within and across years; between individual) at the same site.  Our expectation was that if our method were useful, song similarity values within individuals should be substantially higher than average song similarity values from across-individual comparisons. Human observers can readily recognize that two songs recorded from the same individual in the same year are nearly identical, and that songs from other individuals differ, sometimes by a small amount, and sometimes radically. To assess the reliability of the differences between within and between-individual song similarity levels, we used boot-strapping to calculate 95% confidence intervals around the difference between within-individual and between-individual song similarity scores within each site (R version 3.4.4, see supplement for code). If song similarity scores from within individuals tended to be higher than scores from between individuals, the mean difference should be consistently positive.

Our automated measure of song similarity was modestly effective at recognizing the similarity between songs produced by the same individual bird within the same breeding season. Within each of the seven sites, the mean song similarity score for comparisons between songs of the same individual produced on different dates in the same year was higher than the mean for between-individual comparisons (Figure 1), although typically by only approximately 0.1 correlation units (Table S6). Despite the relatively small average differences, the 95% CIs of these differences were narrow in proportion to the size of the differences and in all but one case, the lower bound was closer to the mean difference than to zero. In no case did the lower bound of the CI encompass zero (Table S6). Thus, we have a consistent and clear, but relatively weak difference in average song similarity scores between within-individual and between-individual song comparisons. In light of this, it was not uncommon for single between-individual scores to be higher than single within-individual scores (Figure 1). Because human observers can readily distinguish songs from the same individual from most other songs produced at a site in a given year, and a human observer would almost never rank two songs from the same individual as less similar than two songs from different individuals, this demonstrates that the automated method produces only modestly reliable measures of song similarity. However, we opted to use the song similarity scores from the automated analyses because quantifying song similarity between all relevant pairs of songs with human judgement would be subjective and prohibitively time consuming due to both the large number of comparisons we wished to make and the difficulty in scoring degree of similarity in songs when those songs differ in many distinct ways. Further, in a preliminary analysis, we experimented with a common form of song similarity quantification known as cross-correlation (Raven 1.3; supplement), and concluded that it was unsuitable. The cross-correlation analyses, while also producing higher values of song similarity on average for within-individual comparisons than across-individual comparisons, produced a much higher range of variability for within-individual comparisons (Figure S2, S3) than did the method we used here (Figure 1).

## Song similarity in biological context

Two of our predictions were that song similarity declines across time and across space. To assess these predictions, within each of our four grassland and three cropland focal sites, we assessed the relationships between song similarity and both (a) the time (in years) between individual recordings and (b) the geographic distance between individual recordings (in meters). Because our measures of song similarity and temporal and geographic distance involved all possible pairwise comparisons within a site, each individual song was included in many song similarity measures. This violates the assumption of parametric statistics that observations are independent. So, instead we used a permutation method that does not make this assumption (R version 3.4.4, see https://(redacted for double blind review) for code). We used MRM (multiple regression on distance matrices) as described by Lichstein (2007) and as can be implemented in the ecodist package (Goslee and Urban 2017) in R. That is, we used linear models in combination with randomized procedures to test the linear association of song-similarities with space and time. In these analyses, the matrix of song-similarities was the dependent matrix, while the matrices of temporal and spatial distances were the independent matrices. MRM randomly permutes the rows and columns of the dependent matrix while the independent matrices remain fixed. We ran one thousand permutations to generate a null distribution of relationships between song similarity scores and both time differences and spatial distances. We could then compare the observed regression coefficients to this null distribution of coefficients to evaluate the strength of the evidence against the null hypothesis that song similarity was unrelated to time differences or spatial distance between recordings. We generated 1-sided p-values because our hypotheses had clear directional predictions. The p-value was the proportion of the regression coefficients from the permutation analysis that were more steeply negative than the observed coefficients. We predicted negative slopes for the relationships between temporal differences and song similarity and between geographic distance and song similarity. Further, we considered the possibility that as time between recordings increased, the geographic distance between those recordings would become a less effective predictor of the similarity between those songs since song features could shift on the landscape across years. To evaluate this, we tested for a positive interaction term between time and distance. We evaluated the importance of this interaction by comparing the change in the $R^2$ value between the model containing only main effects and the model with the interaction term. We used the lm function in base R (version 3.4.4, (R_Core_Team 2019) to generate parameter estimates for each permutation using the methods underpinning the ecodist package (Goslee and Urban 2017) as a model. We did not use the ecodist package itself because that package did not allow us to assess the interaction models, nor could it produce 1-tailed p-values.

Two other predictions of our hypotheses are that song similarity should decline more steeply over time and over distance in locations where site fidelity is lower (cropland, see results) than in locations where site fidelity is higher (grassland, see results). The analyses described above provided slopes and p-values for the relationships between time and song similarity and between distance and song similarity within each of our seven sites. This allowed us to assess our prediction qualitatively by simply asking if the slopes were steeper for the three cropland sites than for the four grassland sites. However, to estimate how likely it was that we would observe a difference in slopes (for song similarity vs. time lag and for song similarity vs. geographic distance) as extreme as this by chance alone, we conducted another round of permutations (R version 3.4.4, see https://(redacted for double blind review) for code). In this analysis, we permuted the song-similarity matrix for each of the seven sites 1000 times, each time calculating the corresponding slope coefficients for each site for time vs. song similarity and for distance vs. song similarity. Then, separately for the time and distance slopes, we calculated the average slope for each of the two location types (cropland and grassland) and then calculated the difference in average slope between the two location types. That is, for each permutation we calculated the slope of the four grassland sites and three croplands sites, took the average of the slopes for the grassland sites and the average of the slopes for the cropland sites, then took the difference between average slopes of cropland and grassland sites. We repeated this process 1000 times. If our hypothesis were correct, we expected our observed difference in average slope between cropland and grassland sites to be greater than most of the simulated differences. We assessed this prediction separately for the time vs. song similarity slopes and the distance vs. song similarity slopes.

In all our statistical analyses we follow recommendations to avoid dichotomous determinations of ‘significance’ and instead to focus on strength of evidence, statistical clarity, and on avoiding over-confidence in results from single studies (Stoehr 1999, Amrhein et al. 2017, Amrhein et al. 2019, Dushoff et al. 2019). This means that we evaluated p-values as continuous evidence bounded by 0, representing “complete incompatibility” between the data and the null hypothesis, and 1, representing “perfect compatibility” with the null hypothesis (Greenland et al. 2016). Small differences in p-values, such as between 0.1 and 0.05 show low repeatability, and thus should not be interpreted as containing meaningfully different information (Amrhein et al. 2017). In cases of relatively large p-values, indicating that the data would not be particularly unlikely given the null, we interpret the observed pattern with caution and withhold strong conclusions until further data can be gathered (Amrhein et al. 2017).

## Results

### Mark-Recapture Analyses

We found strong support for a difference between our grassland and cropland habitats in all three models with ΔAICc < 2 for both apparent survival and detection probability (3hab; Table 1). This variable (3hab) also reflects the influence of differences in field methods between the Author 1 and Author 2 grassland data sets on estimates of apparent survival and detection probability. Ignoring the differences in field methods between the Author 1 and Author 2 data sets by combining the apparent survival and detection probability for both grassland datasets (2hab) had much less support, as that variable did not occur in any of the six highest-ranked models (Table 1). We also found strong support for a difference in apparent survival within a breeding season versus over the non-breeding season (ssn), as this variable was present in all three models with ΔAICc < 2 (Table 1). Further, we considered the possibility of a difference in detection probability between the first and subsequent survey rounds within a breeding season in the Author 1 dataset (rnd; included because of differences in observers between first and later survey rounds). Evidence for the inclusion of such an effect and whether it was similar across habitat types was mixed, as the variable was included in only two of the three highest-ranked models (Table 1). To control for this model uncertainty, we model-averaged all our estimates.

Estimated rates of apparent survival were substantially lower on cropland sites than on grassland sites when examining the Author 1 data (with standardized methods across grassland and cropland sites). Within a breeding season, dickcissels that bred in croplands had a 19.2% (95% CI = 10.4 – 27.9%) chance of disappearing (dying or emigrating) in the 14-day period between survey rounds. However, birds breeding in grassland had only a 7.2% (Author 1; 95% CI = 3.7 – 10.7%) and 8.7% (Author 2; 95% CI = 4.1 – 13.3%) chance of disappearing over a 14 day period (Table 2). Over the 310 day period between breeding seasons, apparent survival estimates had larger confidence intervals that broadly overlapped between grassland and cropland sites (Table 2). When assessed across the entire year (Table 2), as is relevant for overall population turnover, we found annual apparent survival for cropland that was less than half of that for grassland for the Author 1 data. Annual apparent survival for grassland from the Author 2 data was intermediate. Daily rates of apparent survival were higher in the non-breeding season than in the breeding season (Table 2), as would be expected if male dickcissels sometimes abandon territories in mid-season in response to local conditions. The detection probability of dickcissels was highest in grassland habitats within the Author 2 dataset (0.928; CI = 0.875 – 0.960) as expected given the more intense survey methods, and was somewhat higher in grassland habitats (0.752; CI = 0.691 - 0.804) than in cropland habitats (0.591; CI = 0.441 - 0.726) within the Author 1 dataset, but with overlapping confidence intervals (Table 3; we report weighted average estimates for all variables included in any candidate model in Table S7).

## Song Similarity Analyses

The median song similarity scores based on within site and within year comparisons, but averaged across all years, were higher at our grassland sites than at our cropland sites (Figure 1). When we instead looked at song similarity scores from individual years separately, in three of five years all grassland sites had higher median song similarity than all cropland sites. In the other two years, one grassland site had a median song similarity value lower than a cropland site, but all the other grassland sites had higher values than all cropland sites (Figure S4). However, we observed considerable variability in song similarity within both types of sites, and interquartile ranges of grassland and cropland sites often overlapped, thus undermining statistical clarity. The intercepts from the linear models predicting song similarity from time and distance were all larger from the grassland sites than the cropland sites in both the model with only main effects of distance and time (Table 4) and the model including an interaction term between distance and time (Table5). These differences are consistent with higher song similarity in grasslands than in croplands within a given year (time 0) and between individuals in close proximity (distance near 0).

All the relationships between song similarity and time within sites had negative slopes (Table 4, 5), though p-values for each site ranged from small ($p_{rand} < 0.001$), indicating clear evidence of the negative slope to moderately large ($p_{rand} = 0.33$), providing much weaker evidence of a negative slope. Although these variable p-values suggest the relationships varied in strength across sites, negative slopes are consistent with song similarity values across years having declined as the number of intervening years increased. The three steepest declines of song similarity over time in the main-effects model were in the three cropland sites (Table 4), and in the interaction model, three of the four steepest declines in song similarity with time were in the three cropland sites (Table 5). The difference between the average slope for cropland sites and the average slope for grassland sites in the main effects model was larger than 88% of these differences when randomized through permutation analyses. This is equivalent to a p-value of 0.12, which should be interpreted in light of our small sample size (n = 7) as only moderate evidence of a more rapid decline in song similarity over time in croplands than in grasslands.

At six of the seven sites, relationships between song similarity and geographic distance had negative slopes (Table 4, 5, $p_{rand}$ for negative slopes ranged from < 0.0001 – 0.260), indicating with varying degrees of confidence that song similarity values declined as the geographic distance between individuals increased even on these relatively small sites. However, we could discern no clear pattern of differences in these slopes between cropland and grassland sites, in part because the slopes varied substantially among sites within habitat types (Table 4, 5). Consistent with this lack of evident pattern, when we conducted a permutation test we found that differences between the average slope for cropland sites and the average slope for grassland sites in the main-effects model was larger than only 30% of the randomized differences. This corresponds to a p-value of 0.70, which is a lack of clear evidence regarding any differences in the geographic decline in song similarity between cropland and grassland.

We lacked consistent evidence for a strong positive interaction between time and distance. For two of the seven sites, the interaction term was negative, counter to prediction. For the five sites with positive interactions, including the interaction term increased the $R^2$ by < 1% over the main effects model in two cases and >5% over the main effects model in only two cases (Table 6).

## Discussion

Our results are consistent with the hypotheses that song similarity in dickcissels declines over time and space and that vocal culture in dickcissels changes more quickly over time in cropland landscapes where population turnover is faster than in grassland where populations are more stable. Within a given breeding season, dickcissels were more than twice as likely to disappear from our cropland sites, where crop harvest in mid-breeding season presumably often destroyed nests, than from our grassland sites. At all sites, similarity between songs declined as the number of years between the recording of those songs increased thus indicating cultural change over time, though some of these trends were more convincing than others. These observed declines add to a growing body of evidence that that vocal culture in songbird populations can change over time (McGregor and Thompson 1988, Holland et al. 1996, Derryberry 2009, Ju et al. 2019). The slopes of these relationships between time and song similarity were steepest for the three cropland sites, indicating fastest cultural change at those locations. The statistical evidence for this pattern was only moderately clear, however, and so further work will be needed to assess it. Song similarity between pairs of dickcissels also declined over geographic distance within 6 of the 7 sites, but there were no obvious trends in the differences in this relationship between cropland and grassland sites, and so it may be that geographic changes in vocal culture are not strongly driven by differences in population turnover, at least at the scale of our sampling. It may also be that the spatial decline in song similarity happens over shorter distances in grasslands than in croplands, counter to our prediction. However, we have modest evidence of lower overall song similarity in cropland than in grassland sites and so population turnover may influence variation in cultural conformity. It is notable that we were able to detect these patterns despite having only a moderately reliable measure of song similarity. Although we have presented novel evidence for a role of reduced site fidelity in croplands driving an increased rate of cultural evolution in the dickcissels in those systems, our samples were limited to seven sites (4 grassland, 3 cropland), and so sampling at a larger number of sites spread more widely on the landscape is an important next step.

Cropland sites, which had much lower site fidelity than the grassland sites, showed evidence of faster change in vocal culture. Although this pattern is intuitive and is consistent with simulation models (Goodfellow and Slater 1986, Nunn et al. 2009), we know of no prior empirical study that has sought to link site fidelity to rates of cultural change. It seems plausible that both importation of new song features and the accumulation of learning errors drove change in our system because we have observed both the appearance of dramatically different song components and slight changes in note duration, frequency, and shape (Figure S5).

Although we found evidence that reduced site fidelity may drive lower levels of song similarity in dickcissels, such an effect may depend on features of song learning and culture that vary among species (Ortega et al. 2014). In the dickcissel, dialect conformity is high, and each male sings a single song. Thus when established males emigrate, this provides opportunities for new songs to enter the local culture via immigration or learning errors. However, in species in which local vocal culture is characterize by diversity rather than conformity, individuals each sing multiple song types, and these different song types are learned from different tutors, a reduction in site fidelity could reduce the diversity of adult tutors, and thus reduce the potential diversity of song types available for copying. This is what Ortega et al. (2014) found in chipping sparrows (*Spizella passerina*), where locations with lower site fidelity had lower diversity of song types. Thus to predict the role of site fidelity on song similarity for any given species, we should know the degree of local conformity, number of song types sung per male, and possibly also the number of tutors typically imitated by males learning songs, especially if males sing multiple song types.

Knowledge of the number of tutors could potentially provide other insights into the development of vocal culture as well. Unfortunately, we do not yet know the details of how dickcissel males learn song in a vocal neighborhood or how many males in a neighborhood tend to converge on a single song type in a stable population. It may be that males prefer to learn the most common song, but that low site fidelity leads simultaneously to low abundance of even the most common songs and to frequent introductions of new songs that are more likely to be copied than they might be if one existing song were already widely sung. To understand why low song similarity is associated with low site fidelity in dickcissels, we should investigate how dickcissels decide which songs to copy.

Although song similarity declined over geographic distance in all four grassland sites and two of three cropland sites, the slopes of these relationships were not consistently steeper in either habitat type. This suggests that low site fidelity may not lead to a faster rate of decline in song similarity over geographic distance, or even that low site fidelity may reduce this decline in song similarity over space. However, the intercept we observed in our models was consistently lower for cropland sites than for grassland sites, and so if slopes of decline do not differ consistently between these two types of sites, at any given distance, song similarity should on average be lower in our cropland sites than in our grassland sites. This said, we should also consider the possibility that the geographic scale at which we sampled within these sites may be too small for an ideal test of this effect. Earlier research on dickcissels observed declines in song similarity even at relatively close distances, though the largest data set showed the steepest declines in song similarity between 1 and 2 km (Schook et al. 2008). However, within our sites for the current study, nearly all individuals were within 1 km of each other. Thus, before we dismiss the hypothesis that reduced site fidelity promotes more rapid decline in song similarity with geographic distance in dickcissels, we should examine song similarity patterns over longer distances, and of course with larger samples.

Habitat features can influence the evolution of avian vocal culture through direct selection on song propagation (Ey and Fischer 2009), but we are reasonably confident that this mechanism is not driving the patterns that we report here. We are confident because in all of our study sites, and across the geographic range of this species, males sing from exposed perches in areas of either uniformly low vegetation or with only widely scattered higher trees, and their songs propagate primarily through the open air with few or no obstructions. Another possible source of selection on vocal culture may be ambient noise (Slabbekoorn and den Boer-Visser 2006), and two of our grassland sites were close enough to a major highway that, in places on these sites, ambient noise may have been somewhat elevated. However, for most of our birds in both cropland and grassland sites, ambient noise did not obviously differ. Thus, neither of the two environmental selection pressures demonstrated to influence song culture evolution in birds appears likely to differ between our site types.

We generated two separate estimates of apparent survival of territorial male dickcissels occupying grassland sites because we based these estimates on two data sets gathered with different methods. These two data sets produced fairly different apparent survival estimates, and we think that this represents differences in the set of individuals who were marked and observed. With one set of methods, researchers (led by Author 1) banded birds beginning approximately two weeks after male dickcissels began settling on territories. These researchers banded birds for approximately two weeks, and then surveyed each site for banded birds once every two weeks, except in seasons where they surveyed sites only once. With the other set of methods, researchers (led by Author 2) banded and surveyed birds nearly daily throughout the season. This latter set of methods often captured birds who did not appear to be long-term territory holders. Early in the season when male dickcissels were first arriving, these briefly-present individuals may have been prospecting for territories during migration stopover. Later in the season, such individuals likely were adults who had abandoned a territory elsewhere and were wandering the landscape, possibly prospecting for territories for the following year. The inclusion of these individuals that never established long-term territories seems likely to have led to the lower estimated rates of apparent survival from the Author 2 data. For this reason, we feel confident that the Author 2 grassland data are not as appropriate as the Author 1 grassland data for comparison to the Author 1 cropland data, where the methods of capture and re-sighting were standardized across grassland and cropland sites. Thus we also feel confident that the relative differences in apparent survival between the grassland and cropland sites calculated from the Author 1 data represent real biological differences in site fidelity between these habitats in the sites we studied.

This work makes an important contribution by calling attention to the possibility that population dynamics may influence the rate of cultural evolution in dickcissels and other taxa. In dickcissels, we would like to see the validity of our findings assessed with a larger sample of sites. We are also looking forward to improvements in our ability to quantify differences between songs of pairs of birds and thus more precise estimates of the strength of the effects of population turnover on vocal culture. Further, as we mention above, exploring how individual colonists choose song tutors should help us understand the mechanistic links between population turnover and cultural change. We also hope that researchers begin to test for relationships between population processes and cultural change across a range of taxa and types of culture. Certainly we hope that bird song researchers investigate this link, but it seems plausible that population turnover could influence change in a wide range of cultural attributes. These questions may be especially relevant given the growing recognition that cultural diversity is a component of biodiversity and is worthy of conservation (Ryan 2006, Laiolo and Tella 2007).

## Acknowledgements

This project would not have been possible without the support of Brett Sandercock and members of his lab over multiple field seasons. Students who worked on this project were supported by Whitman College and/or by Kansas State University Biology NSF-REU grants (DBI – 0552930 and 0851835). We are grateful to the Konza Prairie Biological Station, V. Schaffer at the Kansas State University Department of Agronomy, and S. Strauss for access to study sites. Students who recorded songs used in this paper but are not authors on this paper include Derek Schook, Carina Castro, Brian Kearns, and Ellie Sterne. T. Parker completed the song processing and many of the analyses while on sabbatical granted by Whitman College. T. Parker thanks K. Nicolaysen for support during this long and time consuming project. S. Edmondson completed the statistical comparisons of song similarity to time and distance while on sabbatical provided by Whitman College. Birds were marked under permit from the USGS and State of Kansas and with the approval of the Whitman College Animal Care and Use Committee. This work complies with the laws of the USA. The authors declare no conflict of interest.

THP and WJ conceived of and designed the study. THP, BS, CF, AS, HK, KB, and ER implemented the field study. THP, STL, SE, CF, MGR, MS, and BV conducted data analyses. THP, SE, CF, AS, HK, KB, ER, BV, and WJ interpreted results. THP wrote the manuscript. THP, STL, SE, HK, MS, BV, WJ edited the manuscript.

data and code are archived at: https://osf.io/em43d/
DOI 10.17605/OSF.IO/EM43D

a pre-print of this manuscript is available at: https://ecoevorxiv.org/vct6e/ https://doi.org/10.32942/osf.io/vct6e

## References

Amrhein, V., F. Korner-Nievergelt, and T. Roth. 2017. The earth is flat (p > 0.05): significance thresholds and the crisis of unreplicable research. PeerJ **5**:e3544.

Amrhein, V., D. Trafimow, and S. Greenland. 2019. Inferential statistics as descriptive statistics: there is no replication crisis if we don’t expect replication. The American Statistician **73**:262-270.

Aplin, L. M. 2019. Culture and cultural evolution in birds: a review of the evidence. Animal Behaviour **147**:179-187.

Baker, M. C., and M. A. Cunningham. 1985. The biology of bird-song dialects. Behavioral and Brain Sciences **8**:85-133.

Baker, M. C., and D. B. Thompson. 1985. Song dialects of White-Crowned Sparrows: historical processes inferred from patterns of geographic variation. Condor **87**:127-141.

Beecher, M. D., and E. A. Brenowitz. 2005. Functional aspects of song learning in songbirds. Trends In Ecology & Evolution **20**:143-149.

Beecher, M. D., and S. E. Campbell. 2005. The role of unshared songs in singing interactions between neighbouring song sparrows. Animal Behaviour **70**:1297-1304.

Bioacoustics_Research_Program. 2011. Raven Pro: Interactive Sound Analysis Software. The Cornell Lab of Ornithology. http://www.birds.cornell.edu/raven., Ithaca, NY.

Bitterbaum, E., and L. F. Baptista. 1979. Geographical variation in songs of California House Finches (*Carpodacus mexicanus*). Auk **96**:462-474.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretical approach. 2nd edition. Springer-Verlag, New York.

Derryberry, E. P. 2009. Ecology shapes birdsong evolution: variation in morphology and habitat explains variation in white-crowned sparrow song. American Naturalist **174**:24-33.

Dushoff, J., M. P. Kain, and B. M. Bolker. 2019. I can see clearly now: reinterpreting statistical significance. Methods in Ecology and Evolution **10**:756-759.

Ey, E., and J. Fischer. 2009. The "acoustic adaptation hypothesis" - a review of the evidence from birds, anurans and mammals. Bioacoustics **19**:21-48.

Fayet, A. L., J. A. Tobias, R. E. Hintzen, and N. Seddon. 2014. Immigration and dispersal are key determinants of cultural diversity in a songbird population. Behavioral Ecology **25**:744-753.

Frawley, B. J., and L. B. Best. 1991. Effects of mowing on breeding bird aundance and species composition in alfalfa fields. Wildlife Society Bulletin **19**:135-142.

Goodfellow, D. J., and P. J. B. Slater. 1986. A model of bird song dialects. Animal Behaviour **34**:1579-1580.

Goslee, S., and D. Urban. 2017. ecodist: Dissimilarity-Based Functions for Ecological Analysis. R package.

Greenland, S., S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman, and D. G. Altman. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. European Journal of Epidemiology **31**:337-350.

Große Ruse, M., D. Hasselquist, B. Hansson, M. Tarka, and M. Sandsten. 2016. Automated analysis of song structure in complex birdsongs. Animal Behaviour **112**:39-51.

Handley, H. G., and D. A. Nelson. 2005. Ecological and phylogenetic effects on song sharing in songbirds. Ethology **111**:221-238.

Holland, J., P. K. McGregor, and C. L. Rowe. 1996. Changes in microgeographic song variation of the Corn Bunting *Miliaria calandra*. Journal Of Avian Biology **27**:47-55.

Igl, L. D. 1991. The role of climate and mowing on Dickcissel (*Spiza americana*) movements, distribution and abundance. Iowa State University, Ames, IA.

Ju, C., P. C. Mundinger, F. C. Geller, and D. C. Lahti. 2019. Four decades of cultural evolution in house finch songs. The Auk: Ornithological Advances **136**.

Krebs, J. R., and D. E. Kroodsma. 1980. Repertoires and geographic variation in bird song. Pages 143-177 *in* J. S. Rosenblatt, R. A. Hinde, C. Beer, and M. C. Busnel, editors. Advances in the Study of Behavior, 11. Academic Press, New York.

Kroodsma, D. E., M. C. Baker, L. F. Baptista, and L. Petrinovich. 1985. Vocal 'dialects' in Nuttall's White-Crowned Sparrow. Current Ornithology **2**:103-133.

Kroodsma, D. E., and F. C. James. 1994. Song variation within and among populations of red-winged blackbirds. Wilson Bulletin **106**:156-162.

Kroodsma, D. E., J. Sanchez, D. W. Stemple, E. Goodwin, M. L. Da Silva, and J. M. E. Vielliard. 1999. Sedentary life style of neotropical sedge wrens promotes song imitation. Animal Behaviour **57**:855-863.

Lachlan, R. F., and P. J. B. Slater. 2003. Song learning by chaffinches: how accurate, and from where? Animal Behaviour **65**:957-969.

Laiolo, P., and J. L. Tella. 2005. Habitat fragmentation affects culture transmission: patterns of song matching in Dupont's lark. Journal Of Applied Ecology **42**:1183-1193.

Laiolo, P., and J. L. Tella. 2007. Erosion of animal cultures in fragmented landscapes. Frontiers In Ecology And The Environment **5**:68-72.

Lemon, R. E. 1975. How birds develop song dialects. Condor **77**:385-406.

Lichstein, J. W. 2007. Multiple regression on distance matrices: a multivariate spatial analysis tool. Plant Ecology **188**:117-131.

McGregor, P. K., and D. B. A. Thompson. 1988. Constancy and change in local dialects of the corn bunting. Ornis Scandinavica **19**:153-159.

Mundinger, P. C. 1975. Song dialects and colonization in the house finch, *Carpodacus mexicanus*, on the east coast. Condor **77**:407-422.

Newton, I. 1998. Population Limitation in Birds. Academic Press, San Diego.

Nunn, C. L., P. H. Thrall, K. Bartz, T. Dasgupta, and C. Boesch. 2009. Do transmission mechanisms or social systems drive cultural dynamics in socially structured populations? Animal Behaviour **77**:1515-1524.

Ortega, Y. K., A. Benson, and E. Greene. 2014. Invasive plant erodes local song diversity in a migratory passerine. Ecology **95**:458-465.

Payne, R. B., L. L. Payne, and S. M. Doehlert. 1988. Biological and cultural success of song memes in indigo buntings. Ecology **69**:104-117.

R_Core_Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ryan, S. J. 2006. The role of culture in conservation planning for small or endangered populations. Conservation Biology **20**:1321-1324.

Sandsten, M., M. Große Ruse, and M. Jönsson. 2016. Robust feature representation for classification of bird song syllables. EURASIP Journal on Advances in Signal Processing **2016**:68.

Schartz, R. L., and J. L. Zimmerman. 1971. The time and energy budget of the male dickcissel (*Spiza americana*) Condor **73**:65-76.

Schook, D. M., M. D. Collins, W. E. Jensen, P. J. Williams, N. E. Bader, and T. H. Parker. 2008. Geographic patterns of song similarity in the dickcissel, *Spiza americana*. Auk **125**:953-964.

Slabbekoorn, H., and A. den Boer-Visser. 2006. Cities change the songs of birds. Current Biology **16**:2326-2331.

Slater, P. J. B. 1986. The cultural transmission of bird song. Trends In Ecology & Evolution **1**:94-97.

Stoehr, A. M. 1999. Are significance thresholds appropriate for the study of animal behaviour? Animal Behaviour **57**:F22-F25.

Temple, S. A. 2002. Dickcissel (*Spiza americana*).*in* A. Poole and F. Gill, editors. The Birds of North America, No. 703. The Birds of North America, Inc., Philadelphia.

Tracy, T. T., and M. C. Baker. 1999. Geographic variation in syllables of House Finch songs. Auk **116**:666-676.

White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study **46 Supplement**:120-138.

Williams, J. M., and P. J. B. Slater. 1990. Modelling bird song dialects - the influence of repertoire size and numbers of neighbors. Journal Of Theoretical Biology **145**:487-496.

Winker, K., J. H. Rappole, and M. A. Ramos. 1995. The use of movement data as an assay of habitat quality. Oecologia **101**:211-216.

Zimmerman, J. L. 1982. Nesting success of dickcissels (Spiza americana) in preferred and less preferred habitats. Auk **99**:292-298.

Zimmerman, J. L., and E. J. Finck. 1989. Philopatry and correlates of territorial fidelity in male dickcissels. North American Bird Bander **14**:83-85.

**TABLE 1.** Cormack-Jolly-Seber (CJS) model selection results for dickcissels captured and resighted in northeastern Kansas, USA, between 2006 and 2011. CJS models estimate apparent survival (ϕ) and probability of capture (*p*). We selected models based on ΔAICc values and Akaike weights ($w_i$). We modeled apparent survival separately for cropland and grassland habitats (2hab), for cropland and the two sources of grassland habitat data (3hab), and for periods within and between breeding seasons (ssn), or we kept apparent survival constant across these conditions (con). Detection probability was modeled separately for cropland and grassland habitats (2hab), for cropland and the two sources of grassland habitat data separately (3hab), for the first survey of each year and subsequent surveys rounds in the Author 1 dataset (rnd), or kept constant (con).

| **Model Structure** | | **Model Statistics** | | | | |
|---|---|---|---|---|---|---|
| **ϕ** | ***p*** | **K** | **Deviance** | **AICc** | **ΔAICc** | **$w_i$** |
| 3hab × ssn | 3hab × rnd | 11 | 302.40 | 1291.46 | 0.00 | 0.331 |
| 3hab × ssn | 3hab + rnd | 10 | 304.66 | 1291.67 | 0.21 | 0.299 |
| 3hab × ssn | 3hab | 9 | 307.10 | 1292.06 | 0.60 | 0.245 |
| 3hab + ssn | 3hab × rnd | 9 | 309.97 | 1294.93 | 3.46 | 0.059 |
| 3hab + ssn | 3hab + rnd | 8 | 312.67 | 1295.59 | 4.13 | 0.042 |
| 3hab + ssn | 3hab | 7 | 315.84 | 1296.72 | 5.26 | 0.024 |
| 2hab × ssn | 3hab × rnd | 9 | 324.17 | 1309.13 | 17.67 | 0.000 |
| 2hab × ssn | 3hab + rnd | 8 | 326.57 | 1309.49 | 18.03 | 0.000 |
| 2hab + ssn | 3hab × rnd | 8 | 326.68 | 1309.59 | 18.13 | 0.000 |
| 2hab × ssn | 3hab | 7 | 328.81 | 1309.69 | 18.23 | 0.000 |
| 2hab + ssn | 3hab + rnd | 7 | 329.63 | 1310.50 | 19.04 | 0.000 |
| 2hab + ssn | 3hab | 6 | 332.27 | 1311.12 | 19.66 | 0.000 |
| 3hab × ssn | 2hab | 8 | 330.04 | 1312.95 | 21.49 | 0.000 |
| ssn | 3hab × rnd | 7 | 333.85 | 1314.73 | 23.27 | 0.000 |
| 3hab | 3hab × rnd | 8 | 333.18 | 1316.09 | 24.63 | 0.000 |
| ssn | 3hab + rnd | 6 | 337.73 | 1316.58 | 25.12 | 0.000 |
| 3hab | 3hab + rnd | 7 | 337.04 | 1317.92 | 26.46 | 0.000 |
| ssn | 3hab | 5 | 341.22 | 1318.03 | 26.57 | 0.000 |
| 3hab | 3hab | 6 | 342.91 | 1321.75 | 30.29 | 0.000 |
| 3hab × ssn | con | 7 | 342.33 | 1323.21 | 31.75 | 0.000 |
| 3hab + ssn | 2hab | 6 | 345.59 | 1324.43 | 32.97 | 0.000 |
| 2hab × ssn | 2hab | 6 | 350.70 | 1329.54 | 38.08 | 0.000 |
| 2hab + ssn | 2hab | 5 | 354.65 | 1331.47 | 40.01 | 0.000 |
| 2hab | 3hab × rnd | 7 | 352.26 | 1333.14 | 41.68 | 0.000 |
| 2hab | 3hab + rnd | 6 | 356.41 | 1335.25 | 43.79 | 0.000 |
| con | 3hab × rnd | 6 | 358.29 | 1337.13 | 45.67 | 0.000 |
| 3hab + ssn | con | 5 | 361.49 | 1338.31 | 46.85 | 0.000 |
| 2hab | 3hab | 5 | 361.82 | 1338.64 | 47.17 | 0.000 |
| ssn | 2hab | 4 | 364.68 | 1339.47 | 48.01 | 0.000 |
| 2hab × ssn | con | 5 | 362.77 | 1339.58 | 48.12 | 0.000 |
| con | 3hab + rnd | 5 | 363.08 | 1339.90 | 48.44 | 0.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| con | 3hab | 4 | 369.51 | 1344.30 | 52.84 | 0.000 |
| 2hab + ssn | con | 4 | 369.89 | 1344.68 | 53.22 | 0.000 |
| 3hab | 2hab | 5 | 369.18 | 1345.99 | 54.53 | 0.000 |
| 2hab | 2hab | 4 | 381.07 | 1355.86 | 64.40 | 0.000 |
| ssn | con | 3 | 386.85 | 1359.62 | 68.16 | 0.000 |
| con | 2hab | 3 | 389.28 | 1362.05 | 70.59 | 0.000 |
| 3hab | con | 4 | 393.99 | 1368.78 | 77.32 | 0.000 |
| 2hab | con | 3 | 405.45 | 1378.22 | 86.76 | 0.000 |
| con | con | 2 | 426.46 | 1397.22 | 105.76 | 0.000 |

**TABLE 2.** Estimates of daily apparent survival (ϕ) and standard errors for dickcissels monitored between 2006 and 2011 in northeastern Kansas, USA, estimated separately for birds breeding in cropland and grassland habitats and for birds monitored by Author 1’s team (A1) and Author 2’s team (A2). We estimated apparent survival for the periods within and between breeding seasons separately, and calculated apparent survival, standard errors, and 95% confidence intervals for a 14-day period between two survey rounds within each breeding season, and for a 310-day period between breeding seasons. These estimates and associated SE and 95%CI are model averages weighted by Akaike weights.

| **Specific Interval** | **Habitat** | **Team** | **ϕ (daily)** | **SE** | **ϕ (period)** | **SE** | **95% CI** |
|---|---|---|---|---|---|---|---|
| Within Breeding Seasons *(14-Day Interval)* | Cropland | A1 | 0.985 | 0.004 | 0.808 | 0.045 | 0.721 – 0.896 |
| | Grassland | A1 | 0.995 | 0.001 | 0.928 | 0.018 | 0.893 – 0.963 |
| | Grassland | A2 | 0.993 | 0.002 | 0.913 | 0.023 | 0.867 – 0.959 |
| Between Breeding Seasons *(310-Day Interval)* | Cropland | A1 | 0.998 | 0.001 | 0.488 | 0.136 | 0.221 – 0.756 |
| | Grassland | A1 | 0.999 | 0.000 | 0.723 | 0.056 | 0.612 – 0.834 |
| | Grassland | A2 | 0.997 | 0.001 | 0.366 | 0.061 | 0.246 – 0.486 |
| Between Years *(365-Day Interval)* | Cropland | A1 | 0.996 | 0.001 | 0.212 | 0.075 | 0.065 – 0.359 |
| | Grassland | A1 | 0.998 | 0.000 | 0.539 | 0.059 | 0.423 – 0.654 |
| | Grassland | A2 | 0.996 | 0.000 | 0.256 | 0.050 | 0.158 – 0.354 |

**TABLE 3.** Estimates of the detection probability (*p*), standard errors and 95% confidence intervals for dickcissels monitored between 2006 and 2011 in northeastern Kansas, USA. We estimated detection probabilities separately for birds breeding in cropland and grassland habitats, for birds monitored by Author 1’s team (A1) and Author 2’s team (A2), and for the first survey of each year and subsequent surveys rounds in the P dataset to correct for a potential observer effect (see methods). These estimates and associated SE and 95%CI are model averages weighted by Akaike weights.

| **Specific Interval** | **Habitat** | **Source** | ***p*** | **SE** | **95% CI** |
|---|---|---|---|---|---|
| 1st Survey of the Year | Cropland | A1 | 0.706 | 0.081 | 0.527 - 0.838 |
| Other Surveys | Cropland | A1 | 0.591 | 0.075 | 0.441 - 0.726 |
| 1st Survey of the Year | Grassland | A1 | 0.785 | 0.045 | 0.685 - 0.860 |
| Other Surveys | Grassland | A1 | 0.752 | 0.029 | 0.691 - 0.804 |
| All Surveys | Grassland | A2 | 0.928 | 0.021 | 0.875 - 0.960 |

**TABLE 4.** The relationships between pairwise song similarity values and geographic distance between pairs of recordings and time difference (in years) between pairs of recordings, along with the overall model $R^2$ value (song similarity ~ geographic distance + time difference). Each estimate is accompanied by a one-tailed p-value.

| **Site** | **intercept** | $p_{rand}$ | **distance slope** | $p_{rand}$ | **time slope** | $p_{rand}$ | **overall $R^2$** | $p_{rand}$ |
|---|---|---|---|---|---|---|---|---|
| grassland: | | | | | | | | |
| R20A-2A | 0.618 | 0.001 | -1.25E-05 | 0.260 | -7.03E-03 | 0.000 | 0.022 | 0.000 |
| 20C | 0.621 | 0.000 | -6.64E-05 | 0.000 | -1.95E-03 | 0.106 | 0.039 | 0.000 |
| 4F | 0.654 | 0.180 | -3.11E-05 | 0.205 | -9.30E-04 | 0.334 | 0.003 | 0.602 |
| C1B | 0.626 | 0.008 | -3.89E-05 | 0.011 | -4.49E-03 | 0.138 | 0.016 | 0.070 |
| cropland: | | | | | | | | |
| Ashland | 0.588 | 0.005 | 1.28E-05 | 0.658 | -1.21E-02 | 0.002 | 0.085 | 0.009 |
| Strauss | 0.552 | 0.062 | -3.26E-05 | 0.133 | -8.49E-03 | 0.117 | 0.025 | 0.205 |
| North | 0.519 | 0.001 | -5.16E-05 | 0.009 | -8.10E-03 | 0.017 | 0.029 | 0.029 |

**TABLE 5.** The relationships between pairwise song similarity values and geographic distance between pairs of recordings and time difference (in years) between pairs of recordings, an interaction term between distance and time, and the overall model $R^2$ value (song similarity ~ geographic distance + time difference + geographic distance*time difference). Each estimate is accompanied by a p-value, and p-values are one-tailed for the distance and time slopes and the distance x time interaction.

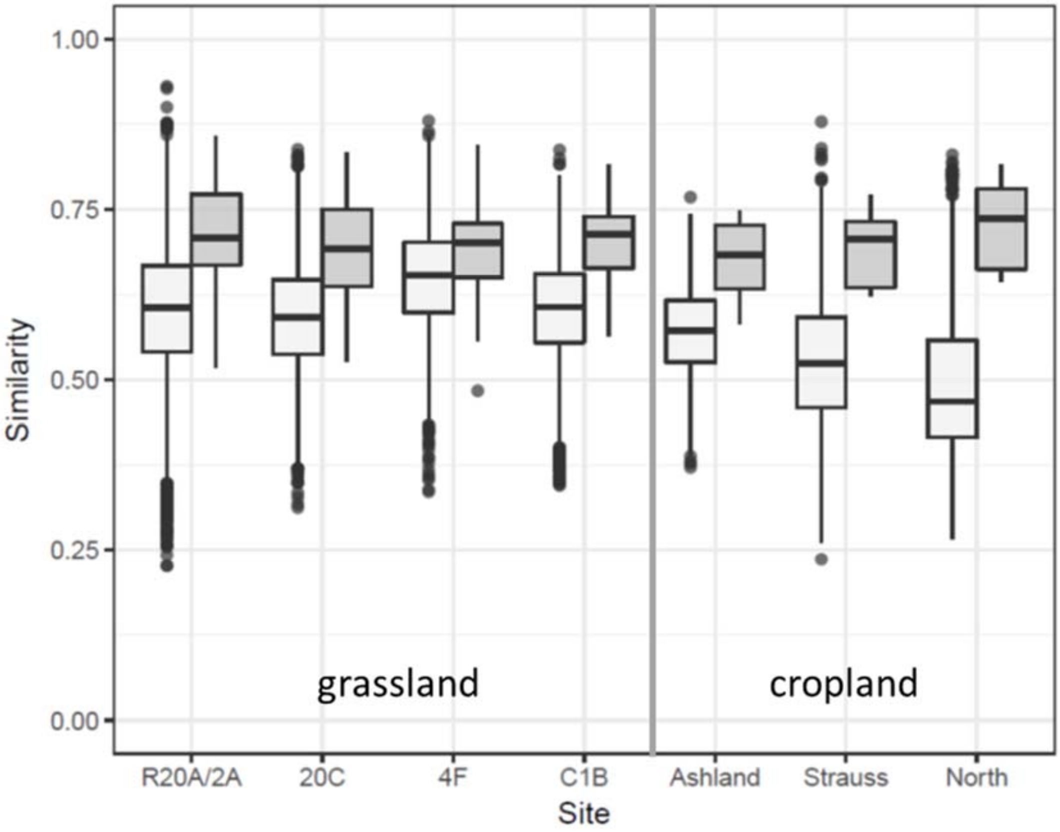
| **Site** | **int.** | $p_{rand}$ | **distance slope** | $p_{rand}$ | **time slope** | $p_{rand}$ | **distance x time** | $p_{rand}$ | **overall $R^2$** | $p_{rand}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| grassland: | | | | | | | | | | |
| R20A-2A | 0.624 | 0.001 | -3.10E-05 | 0.076 | -9.89E-03 | 0.000 | 8.00E-06 | 0.032 | 0.023 | 0.000 |
| 20C | 0.621 | 0.000 | -6.80E-05 | 0.000 | -2.32E-03 | 0.180 | 1.00E-06 | 0.406 | 0.039 | 0.000 |
| 4F | 0.655 | 0.193 | -3.30E-05 | 0.208 | -1.14E-03 | 0.354 | 1.00E-06 | 0.467 | 0.003 | 0.626 |
| C1B | 0.622 | 0.022 | -2.70E-05 | 0.049 | -1.76E-03 | 0.292 | -8.00E-06 | 0.896 | 0.017 | 0.074 |
| cropland: | | | | | | | | | | |
| Ashland | 0.580 | 0.193 | 3.90E-05 | 0.878 | -7.12E-03 | 0.117 | -1.40E-05 | 0.864 | 0.090 | 0.010 |
| Strauss | 0.555 | 0.098 | -4.00E-05 | 0.162 | -1.06E-02 | 0.131 | 5.00E-06 | 0.333 | 0.025 | 0.244 |
| North | 0.527 | 0.000 | -7.50E-05 | 0.003 | -1.27E-02 | 0.005 | 1.30E-05 | 0.030 | 0.031 | 0.005 |

**TABLE 6.** Comparison of the model fit, as measured by $R^2$, between models that use only main effects (geographic distance between pairs of recordings, time difference [in years] between pairs of recordings) to explain pairwise song similarity and models that use both these main effects and an interaction between the two to explain pairwise song similarity.

| Site | main effects $R^2$ | interaction $R^2$ | absolute increase in $R^2$ | percent increase in $R^2$ |
|---|---|---|---|---|
| grassland: | | | | |
| R20A-2A | 0.02206 | 0.02295 | 0.00088 | 4.01 % |
| 20C | 0.03941 | 0.03943 | 0.00002 | 0.04 % |
| 4F | 0.00262 | 0.00262 | 0.00000 | 0.19 % |
| C1B | 0.01652 | 0.01731 | 0.00080 | 4.83 % |
| cropland: | | | | |
| Ashland | 0.08547 | 0.08985 | 0.00438 | 5.12 % |
| Strauss | 0.02504 | 0.02537 | 0.00033 | 1.32 % |
| North | 0.02889 | 0.03109 | 0.00220 | 7.63 % |

Figure headings

**Figure 1.** Plots of median song similarity values for comparisons of recordings of the same individual on different days within a given year (dark grey boxes) at each of the seven sites and for comparisons of recordings of songs of different individuals at these same sites within years (light gray boxes). The central line is the median and the box represents the interquartile range (middle 50th percentile).



## Supplementary materials

### A. Explanation of dickcissel song structure and song learning

The dickcissels is a migrant songbird breeding in grasslands of the Eastern Great Plains of North America (Temple 2002) and wintering primarily in Northern South America. The vast majority of dickcissel males produce a single song (Figure 1), at least while occupying the same breeding territory within and between seasons (Schook et al. 2008; T. Parker unpublished data). All individual's song consist of at least two of three phrase types (Figure S1). One of these three phrase types, the introductory “dick” phrase, is sung by all males. An individual male produces either one or two versions of this “dick” phrase, and each version can be repeated one to several time per song. A series of dick phrases is sometimes preceded by a long series of dick-like phrases that differ in form from the typical dick phrases sung by the male in question. These series appear to be sung by any given individual only rarely, possibly when agitated. We excluded these putative agitation phrases from our analyses because they are not a consistent component of a male's song. Many, but not all, dickcissels include the “cissel” phrase type in their song. Cissels are longer and more complex and contain some higher frequency components. Most commonly a male sings a single version of the “cissel” phrase, and produces this phrase three times in a given song. However, some males produce two versions of the cissel phrase, and an individual version of a cissel phrase can be sung anywhere between one and four times within a song. Some dickcissels sing a “trill” phrase type of rapidly repeated notes. In a trill, the pause duration between repeated sets of notes no longer than the pauses between notes within a set. Some trills have a short set of introductory notes prior to the repeated series. A full trill series is rarely repeated. In contrast, cissels occasionally contain a series of notes that might qualify as a trill, but because cissel phrases have substantial proportions of non-trilled notes and the full phrase is usually repeated two to four times in a predictable pattern, cissel phrases and trill phrases cannot be mistaken for each other. Among dickcissels that produce a trill phrase type, most individual sing only a single version of the trill in their song, but rarely a male can produce two different versions of a trill within a song. Usually only one trill is produced per song. The versions of the phrases produced by each male is highly repeatable during a single breeding season. Further dickcissels show striking geographic structure in their song culture, with most males in close proximity to each other singing almost identical song phrases, and the degree of similarity of phrase types often declining considerably as the distance between recording sites approaches 2 km (Schook et al. 2008). Individual dickcissels' songs also differ from each other in the number of times phrases are repeated and the cadence of these repetitions (Figure 1). However, unlike the content of the phrases, some of the differences in the repetition of phrases are not consistent within and across individuals, and so they have not been the focus of our research (Schook et al. 2008, Parra et al. 2017).

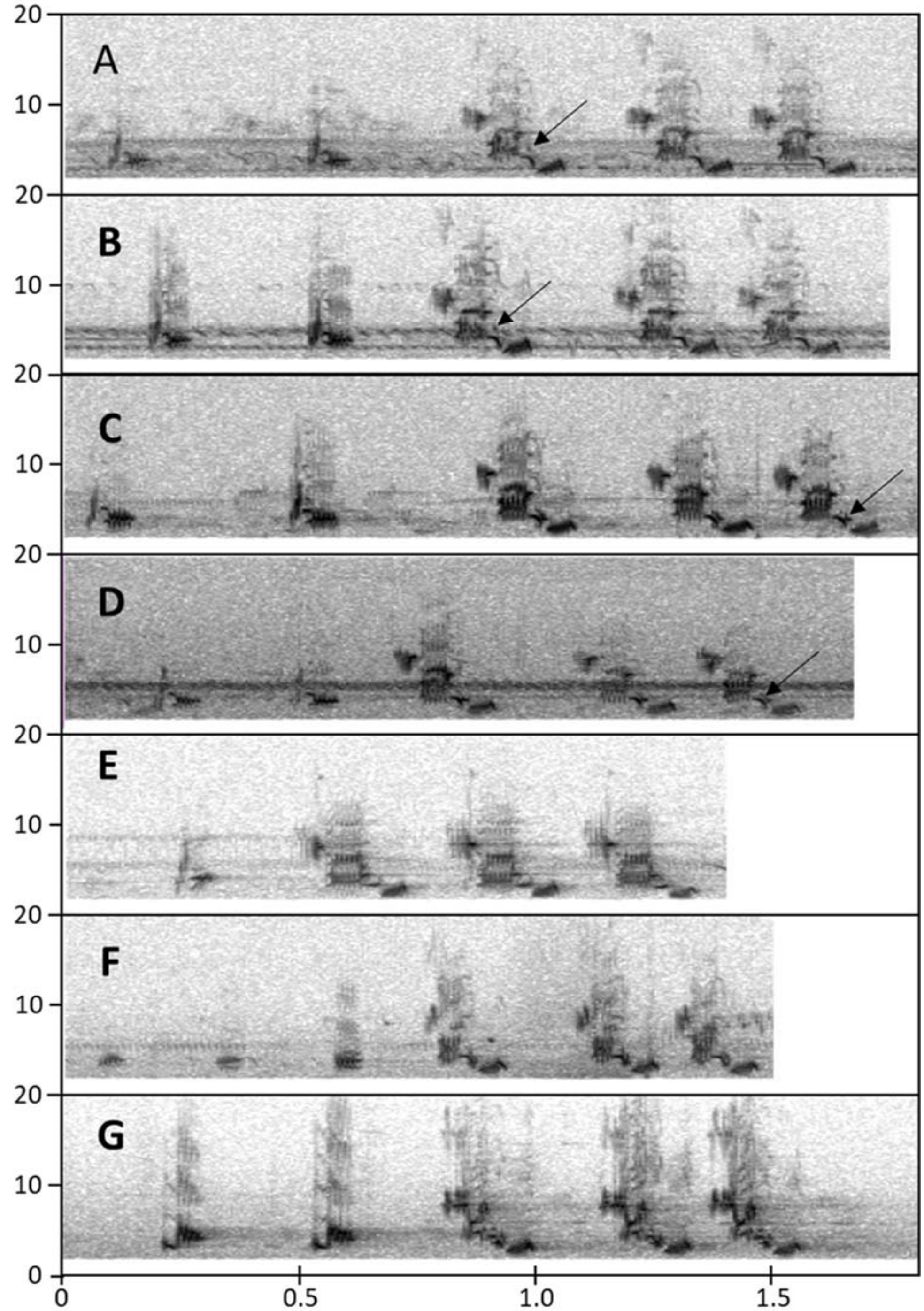
Circumstantial evidence suggests that male dickcissels learn their song from immediate adult neighbors after dispersal (Krebs and Kroodsma 1980). First, song similarity by dickcissels declines with increasing distance on the scale of 100's of meters to several km (Schook et al. 2008), and so pre-migration song learning during the first summer of life would require that, after migration, dickcissels return to a small area within a few hundred meters of where they learned their song(s). However, when nestling dickcissels are banded, they are almost never re-sighted near the location of banding in subsequent years (Temple 2002, Small et al. 2012; B. Sousa unpublished data), indicating that they recruit locally only rarely. Further, because these study sites where nestlings have been banded are often much larger than the extent of a local song neighborhood, we cannot explain this lack of recaptures by assuming that the young birds have dispersed within the song neighborhood but outside the bounds of the study site. If they disperse outside of the study area, they are dispersing into other cultural areas. It also appears unlikely that recently fledged dickcissels disperse into other song neighborhoods beyond the scale of the typical study area to learn a variety of songs prior to migration. This is because fledgling dickcissels typically remain within <1 km of their nest in the month after fledging (Verheijen 2017). Taken together, this evidence suggests that young adult male dickcissels are able to match the songs of their adjacent neighbors only by learning their song when they settle on their first breeding territory after their first winter in South America. An alternative scenario would require males to learn songs during migration or in winter and then to follow a tutor so as to settle near him on the breeding ground, but this seems unlikely, in part because dickcissels migrate at night in flocks of hundreds or thousands of individuals (Temple 2002).

We do not know the extent to which older males learn new songs after having learned the original song in their first breeding neighborhood. We do know that a small proportion of male dickcissels in most locations do not conform to the local culture, and this lack of conformity is often stark, with individuals sometimes producing songs that radically differ from their neighbors. We hypothesize that these non-conformers are males who learned their song elsewhere and then later immigrated into this new vocal culture without adopting the local culture. This is somewhat surprising given that we have observed at least one long-lived individual partially change its song (by adding a new version of the cissel phrase) when the neighborhood culture changed across years and a new cissel phrase version become common. Thus dickcissels seem able to alter their song after their first breeding season, but they may often not do so.

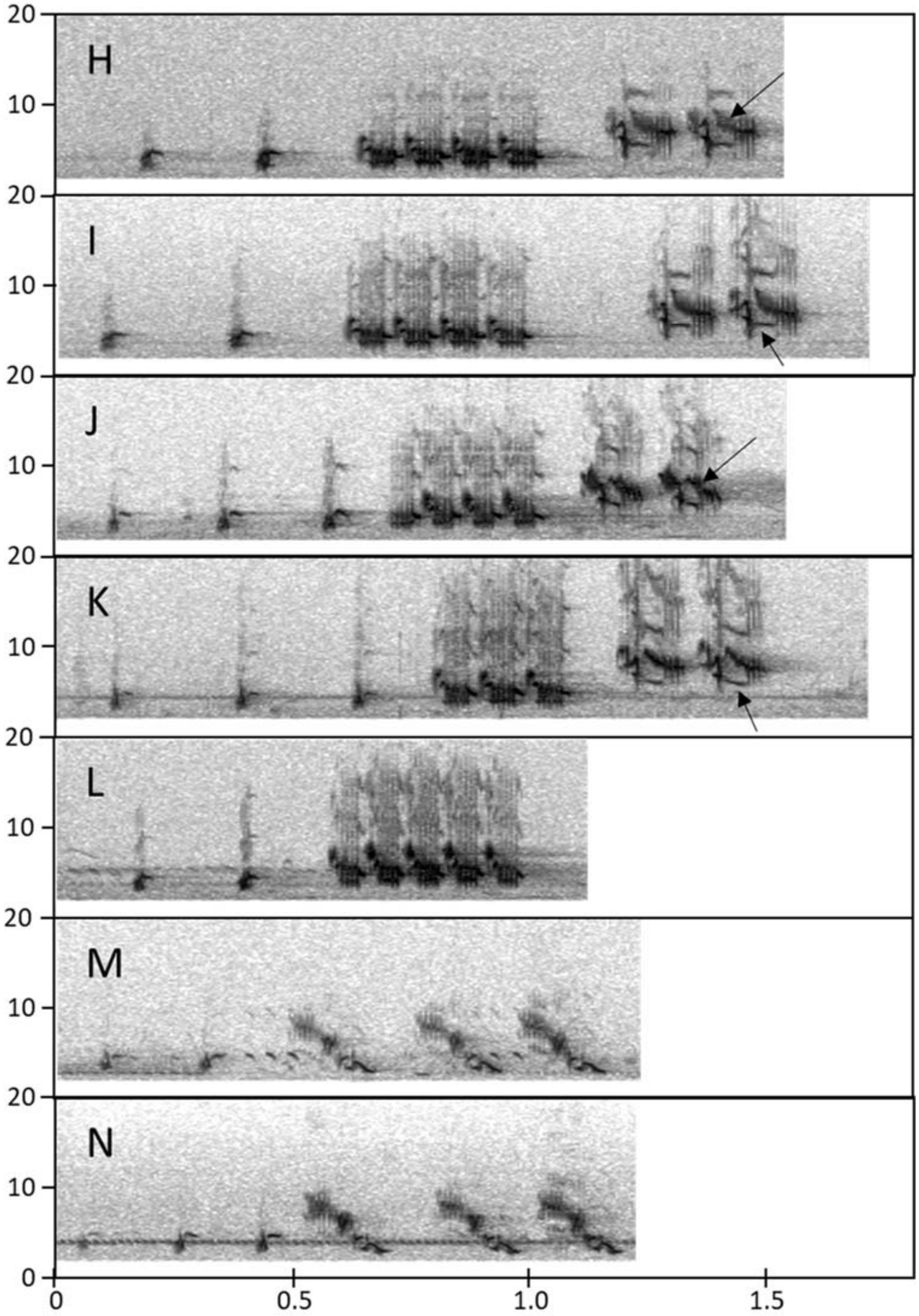
**Figure S1.** Complete songs of multiple male dickcissels from two sites (C1B, a grassland site; North Farm, a cropland site) in 2008, with frequency (kHz) on y-axis and time (seconds) the x-axis. Panels A – G and M and N all depict songs with two phrase types: one brief ‘dick’ phrase often repeated several times followed by several (in this case, always three) very similar ‘cissel’ phrases. The songs in panels H – K also begin with ‘dick’ phrases, but then are followed by a ‘trill’ phrase and then by two ‘cissel’ phrases. Panels A and B depict songs produced by the same male on two different dates, as do panels C and D, H and I, and M and N, with each of these pairs of panels demonstrating the high degree of within-male song consistency. We can also see similarity between the songs of different males within a site, for instance the dick phrases in panels A – E and the cissel phrases in A – F are all very similar as are the dick phrases in panels H – N, the cissel phrases in H – K, and the trill phrases in H – L. However, even very similar phrases differ in consistent ways between individual birds. For instance, the arrows in panels and A and B identify a small note that is shorter and of higher frequency than the corresponding note identified by the arrows in panels C and D. Similarly, the arrows in panels H and J indicate the final series of notes in the cissel phrase which show differences in the pattern of frequency modulation and duration. The arrow in panel I indicates a pure-tone note which remains on a single frequency, while the corresponding note in panel K descends in frequency. Also in panel J, notice that the trill has an introductory portion that differs from the trills in panels H, I, K, and L.

Figure S1

C1B



North



B. Explanation of the origins of this study

Our interests in dickcissel song and our methods of study have evolved over time. We began recording dickcissel song in 2005 and 2006 to assess the geographic extent and pattern of song similarity (Schook et al. 2008). We were also curious about within-individual variation in song, and so in 2006 began color-banding individuals and returning to record individual birds within and across seasons. We continued recording marked birds in 2007 to explore the possibility of change in song across years. In 2008, we initiated the project that is the subject of this paper to study whether changes in population turnover would be associated with the rate of temporal or spatial change in song culture, and so we began a larger effort to mark birds at our existing grassland sites and at a cropland site. In both 2009 and 2010, we added a new cropland site and continued marking birds at all of our sites. We returned to record songs, but not to mark new birds, at most sites in 2011 and 2014. A parallel study recorded song and banded dickcissels at several grassland sites as part of an exploration of sexual selection in this species in 2006-2009 (Sousa and Westneat 2013b, a), and we include these recordings in the analyses for this current paper.

C. Methods – Site locations and sampling

**Table S1.** Locations and sizes of study sites. Sites at Konza Prairie Biological Station (KPBS) are KPBS designated ‘watersheds’ and the number in the name refers to the number of years between controlled early spring burns (prior to settlement by dickcissels). We did not necessarily incorporate an entire watershed into the study, and the area measurement we present represents the portion of the watershed(s) we used. The sites at KPBS averaged 2.3 km apart (when measured between the closest edges) or 2.6 km apart (when measured between site centroids). The closest two sites were 421 m apart at their closest edges. The next closest two sites were 1.7 km apart at their closest edges.

| Site | type | latitude | longitude | approx. area (ha)# |
|---|---|---|---|---|
| KPBS | | | | |
| R20A / 2A | grass | 39.06 | -96.60 | 42 |
| 20C / 1D* | grass | 39.06 | -96.56 | 27 |
| 4F | grass | 39.07 | -96.57 | 16 |
| C1B / C4A | grass | 39.09 | -96.54 | 37 |
| North | crop | 39.21 | -96.59 | 44 |
| Ashland | crop | 39.12 | -96.61 | 25 |
| Strauss | crop | 39.39 | -96.85 | 18 |

# These areas do not match the area of the KPBS designated watersheds because in most cases we did not include entire watersheds in our study.

* All song recordings but one come from 20C, but birds banded and monitored for the mark-recapture study by B. Sousa were in both 20C and adjacent 1D.

**Table S2.** Site by site summary of sampling. We did not visit all sites in all years. Number of visits varied among years. Sampling carried out by T. Parker and students identified by ‘P’. Sampling carried out by B. Sousa and associates identified by ‘S’. Song recording methods were similar between these two groups, though recording equipment differed. Monitoring protocols for mark/re-sighting differed substantially between P and S teams, and thus we considered these data separately for analyses. In 2008 and 2009 both P and S teams monitored birds in 20C. For any birds monitored by both T. Parker and B. Sousa teams, we included only the T. Parker data in the mark-recapture analysis so that we could generate separate mark-recapture estimates for T. Parker and B. Sousa data.

| Site | | year | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 2006 | 2007 | 2008 | 2009 | 2010# | 2011 | 2014* |
| R20A / 2A | visits | 4 | 5 | 1 | 0 | 4 | 1 | 1 |
| | mark/re-sight | S | S | S | | P | P | P |
| | song recording | P, S | S | S | | P | P | P |
| 20C / 1D | visits | 1 | 1 | 3 (1) | 5 (5) | 4 | 1 | 0 |
| | mark/re-sight | P | P | P (S) | P (S) | P | P | |
| | song recording | P | P | P, S | P, S | P | P | |
| 4F | visits | 1 | 0 | 3 | 4 | 4 | 1 | 0 |
| | mark/re-sight | P | | P | P | P | P | |
| | song recording | P | | P | P | P | P | |
| C1B | visits | 0 | 0 | 4 | 4 | 4 | 1 | 1 |
| | mark/re-sight | | | P | P | P | P | |
| | song recording | | | P | P | P | P | P |
| North | visits | 0 | 0 | 4 | 5 | 4 | 1 | 1 |
| | mark/re-sight | | | P | P | P | P | |
| | song recording | | | P | P | P | P | P |
| Ashland | visits | 0 | 0 | 0 | 5 | 4 | 1 | 1 |
| | mark/re-sight | | | | P | P | P | |
| | song recording | | | | P | P | P | P |
| Strauss | visits | 0 | 0 | 0 | 0 | 4 | 1 | 1 |
| | mark/re-sight | | | | | P | P | |
| | song recording | | | | | P | P | P |

# last year birds were banded

* data not used for mark-recapture analysis due to the lack of sampling over prior 2 years and the 4 years since last birds were banded.

**Table S3.** Number of songs analyzed per year for each site. Within a given year, each song was recorded from a different individual male dickcissel.

| Site | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2014 |
|---|---|---|---|---|---|---|---|
| 2A-R20A | 32 | 40 | 15 | 0 | 25 | 33 | 5 |
| 4F | 15 | 0 | 9 | 12 | 11 | 10 | 0 |
| 20C | 26 | 3 | 10 | 19 | 14 | 9 | 0 |
| C1B | 0 | 0 | 14 | 25 | 25 | 18 | 5 |
| Ashland | 0 | 0 | 0 | 11 | 5 | 6 | 5 |
| North | 0 | 0 | 17 | 16 | 10 | 9 | 7 |
| Strauss | 0 | 0 | 0 | 0 | 15 | 10 | 5 |

**Table S4.** Number of individuals from which two songs, each from a different date, were analyzed to assess the sensitivity of our song comparison method.

| Site | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2014 |
|---|---|---|---|---|---|---|---|
| grassland: | | | | | | | |
| 2A-R20A | 12 | 0 | 0 | 0 | 14 | 0 | 0 |
| 4F | 4 | 0 | 7 | 6 | 6 | 0 | 0 |
| 20C | 6 | 0 | 3 | 10 | 6 | 0 | 0 |
| C1B | 0 | 0 | 0 | 4 | 2 | 0 | 0 |
| cropland: | | | | | | | |
| Ashland | 0 | 0 | 6 | 5 | 9 | 0 | 0 |
| North | 0 | 0 | 6 | 6 | 2 | 0 | 0 |
| Strauss | 0 | 0 | 0 | 0 | 5 | 0 | 0 |

D. Methods - Additional details regarding calculation of song similarity scores

Our scheme of classifying songs as dicks, cissels, and trills works well, but it is not perfect. It appears, for instance, that short dick phrases, which are sometimes sung in rapid series, may evolve over time into trills (T. Parker, unpublished observations), and thus there are phrases which could be classified into either category. However, of the many hundreds of songs we have processed and classified, the vast majority required no subjective judgement regarding classification. In nearly all cases, we could adopt a simple decision rule to ensure consistency. For instance, in cases where dick phrases could have been judged to be either dicks or part of a trill series, we considered them dicks if they matched other phrases at the site that were unambiguously classified as dicks.

When comparing song phrases with our quantitative methods, if males at a site sang two versions of the same phrase types, e.g., two distinct dick phrases, we classified the phrases of that type as 'one' and 'two', e.g., 'dick1' and 'dick2', and compared all 'one' phrases to each other and all 'two' phrases to each other but not the 'one' phrases to the 'two' phrases. If two males at a site sang the same two versions of a phrase, they always sang them in the same order, so 'one' and 'two' were not arbitrary labels but reflected the order in which the two versions appeared. If a male sang one commonly shared version and one rare version or locally unique version, we assigned the commonly shared version to category 'one' or 'two' to match the position in which that phrase version was typically sung at the site so that we could compare like to like. If some males sang two versions of a phrase type and other males sang only one version of that phrase type, for the males who sang only one version, we saved two copies of that version, one as version one and the other as version two. This allowed us to assess the quantitative similarity of the given phrase type between males even if they sang different numbers of versions. If a male singing only one version of a phrase type sang that version >1 time, we typically used different renditions of that version as version one and two. The benefit of this was that when comparing two males who both sang only one version of a phrase type, but for whom we had saved two copies of that version, we could increase the precision of our song similarity estimate by basing it on an average of the comparison between 'version one' and 'version two', which in this case were just two renditions of the same version.

We limited our analyses to considering two versions of a given phrase type per male because this appeared to be the best way to minimize bias in the calculation of average similarity. First, it was extremely rare for males to sing what might be interpreted as >2 versions of a given phrase type so there would have been very few cases where including a third phrase type would have provided any additional information. Second, in the very few cases where it would have been possible to classify >2 versions, differences between two of the versions were quantitative rather than qualitative and relatively small. Third, in the much more frequent case when males sang two versions, we could match the versions between two males and so compare like versions of a phrase when quantifying song similarity. Fourth, if two males did not sing the same phrase version, we could quantify how different those two versions were from each other. Further, for all pairs of males regardless of whether either male sang one version or two versions of the phrase type, the average similarity based on two comparisons (one of phrase version 1, the other of phrase version 2) always provided an unbiased estimate of the similarity between the phrases. If instead we had allowed for a third version to account for the very few males who could be considered to sing a third version of a given phrase type, this would have complicated the calculation of averages for several reasons. For a male who sang two versions, we would have had to choose one of those versions to compare with the third version of males who sang three, and this choice would have been arbitrary. When comparing a male who sang two versions of a phrase with either a one-version or a two-version male, one of the versions would have contributed to

2/3 of the weight of the average because it would have been used twice. The only ways around this added substantial complexity to already complex data management and analysis methods. Because this complexity addressed a very rare issue, and a relatively unimportant one where present, we opted for the solution just described.

Before we averaged the two quantitative measures (spectrogram and ambiguity spectrum) of similarity for each phrase type, we wanted to weight the information from both of these quantitative measures identically, and so each measure needed to be on the same scale. Although both of these methods create values that have the potential to range between 0 and 1 (with 0 = identical and 1 = complete lack of similarity; which we later reversed such that 1 = identical, see below), the variability of the filtered ambiguity spectrum method is in practice much lower in our data set. Because we wanted the two measures of similarity to be weighted as equally as possible, we standardized the two matrices to actually vary between 0 and 1 by dividing the value in each matrix cell by the highest value for that measurement type from across all comparisons. This was approximately equal to 1 for the multi-tapered spectrogram data (and so its transformed range ≈ original range), but was 0.308 for the filtered ambiguity spectrum method. After the measurements were standardized, we averaged the values across the two measures. We created this average within each phrase type, or if two versions of a phrase type occurred at a site, within each version of each phrase type. We weighted each phrase type equally when averaging across types. If there were two versions of a phrase type at a given site, each was weighted 0.5 to continue to give equal weight to each broadly defined phrase type. Finally, to reverse the direction of the song similarity score so that it ranged from 0 (no overlap) to 1 (songs identical), we subtracted the song similarity averages from 1.

We generated raw pairwise song phrase similarity scores using MATLAB (version R2016b; see https://osf.io/em43d/ for code).

We combined similarity scores across phrase types and across song comparison methods using R (version 3.3.3 and 3.4.2, see osf.io for code)

## E. Sensitivity of mark-recapture models to possible over-dispersion

Table S5. To assess the robustness of our apparent survival model ranking to possible over-dispersion, we examined model rankings with $\hat{c}$ set to whole numbers between 1 and 6. Note that the variable 3hab (which estimated different values for grassland sites monitored by T. Parker and those monitored by B. Sousa, and cropland sites) remained among the top models (ΔAICc < 2) for both apparent survival and recapture probability in analyses with $\hat{c}$ = 1, 2, 3, and 4. Only when $\hat{c}$ reached 5 did the top models no longer all include an effect of habitat on apparent survival, although this habitat effect remained in the set with Delta QAIC < 2 for both $\hat{c}$ = 5 and 6. Thus our model is robust to over dispersion.

| **Model Structure** | | **Model Statistics** | | | | |
|---|---|---|---|---|---|---|
| **ϕ** | ***p*** | **K** | **Deviance** | **AICc** | **ΔAICc** | $w_i$ |
| $\hat{c}$ = 1 | | | | | | |
| 3hab × ssn | 3hab × rnd | 11 | 302.40 | 1291.46 | 0.00 | 0.331 |
| 3hab × ssn | 3hab + rnd | 10 | 304.66 | 1291.67 | 0.21 | 0.299 |
| 3hab × ssn | 3hab | 9 | 307.10 | 1292.06 | 0.60 | 0.245 |
| 3hab + ssn | 3hab × rnd | 9 | 309.97 | 1294.93 | 3.46 | 0.059 |
| 3hab + ssn | 3hab + rnd | 8 | 312.67 | 1295.59 | 4.13 | 0.042 |
| 3hab + ssn | 3hab | 7 | 315.84 | 1296.72 | 5.26 | 0.024 |
| 2hab × ssn | 3hab × rnd | 9 | 324.17 | 1309.13 | 17.67 | 0.000 |
| 2hab × ssn | 3hab + rnd | 8 | 326.57 | 1309.49 | 18.03 | 0.000 |
| 2hab + ssn | 3hab × rnd | 8 | 326.68 | 1309.59 | 18.13 | 0.000 |
| 2hab × ssn | 3hab | 7 | 328.81 | 1309.69 | 18.23 | 0.000 |
| 2hab + ssn | 3hab + rnd | 7 | 329.63 | 1310.50 | 19.04 | 0.000 |
| 2hab + ssn | 3hab | 6 | 332.27 | 1311.12 | 19.66 | 0.000 |
| 3hab × ssn | 2hab | 8 | 330.04 | 1312.95 | 21.49 | 0.000 |
| ssn | 3hab × rnd | 7 | 333.85 | 1314.73 | 23.27 | 0.000 |
| 3hab | 3hab × rnd | 8 | 333.18 | 1316.09 | 24.63 | 0.000 |
| ssn | 3hab + rnd | 6 | 337.73 | 1316.58 | 25.12 | 0.000 |
| 3hab | 3hab + rnd | 7 | 337.04 | 1317.92 | 26.46 | 0.000 |
| ssn | 3hab | 5 | 341.22 | 1318.03 | 26.57 | 0.000 |
| 3hab | 3hab | 6 | 342.91 | 1321.75 | 30.29 | 0.000 |
| 3hab × ssn | con | 7 | 342.33 | 1323.21 | 31.75 | 0.000 |
| 3hab + ssn | 2hab | 6 | 345.59 | 1324.43 | 32.97 | 0.000 |
| 2hab × ssn | 2hab | 6 | 350.70 | 1329.54 | 38.08 | 0.000 |
| 2hab + ssn | 2hab | 5 | 354.65 | 1331.47 | 40.01 | 0.000 |
| 2hab | 3hab × rnd | 7 | 352.26 | 1333.14 | 41.68 | 0.000 |
| 2hab | 3hab + rnd | 6 | 356.41 | 1335.25 | 43.79 | 0.000 |
| con | 3hab × rnd | 6 | 358.29 | 1337.13 | 45.67 | 0.000 |
| 3hab + ssn | con | 5 | 361.49 | 1338.31 | 46.85 | 0.000 |
| 2hab | 3hab | 5 | 361.82 | 1338.64 | 47.17 | 0.000 |
| ssn | 2hab | 4 | 364.68 | 1339.47 | 48.01 | 0.000 |
| 2hab × ssn | con | 5 | 362.77 | 1339.58 | 48.12 | 0.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| con | 3hab + rnd | 5 | 363.08 | 1339.90 | 48.44 | 0.000 |
| con | 3hab | 4 | 369.51 | 1344.30 | 52.84 | 0.000 |
| 2hab + ssn | con | 4 | 369.89 | 1344.68 | 53.22 | 0.000 |
| 3hab | 2hab | 5 | 369.18 | 1345.99 | 54.53 | 0.000 |
| 2hab | 2hab | 4 | 381.07 | 1355.86 | 64.40 | 0.000 |
| ssn | con | 3 | 386.85 | 1359.62 | 68.16 | 0.000 |
| con | 2hab | 3 | 389.28 | 1362.05 | 70.59 | 0.000 |
| 3hab | con | 4 | 393.99 | 1368.78 | 77.32 | 0.000 |
| 2hab | con | 3 | 405.45 | 1378.22 | 86.76 | 0.000 |
| con | con | 2 | 426.46 | 1397.22 | 105.76 | 0.000 |
| ĉ = 2 | | | | | | |
| 3hab × ssn | 3hab | 9 | 153.55 | 655.14 | 0.00 | 0.231 |
| 3hab + ssn | 3hab | 7 | 157.92 | 655.43 | 0.29 | 0.200 |
| 3hab + ssn | 3hab + rnd | 8 | 156.34 | 655.88 | 0.74 | 0.159 |
| 3hab × ssn | 3hab + rnd | 10 | 152.33 | 655.97 | 0.83 | 0.153 |
| 3hab + ssn | 3hab × rnd | 9 | 154.98 | 656.57 | 1.43 | 0.113 |
| 3hab × ssn | 3hab × rnd | 11 | 151.20 | 656.89 | 1.75 | 0.096 |
| 2hab + ssn | 3hab | 6 | 166.14 | 661.61 | 6.47 | 0.009 |
| 2hab × ssn | 3hab | 7 | 164.41 | 661.91 | 6.77 | 0.008 |
| 2hab + ssn | 3hab + rnd | 7 | 164.81 | 662.32 | 7.18 | 0.006 |
| 2hab × ssn | 3hab + rnd | 8 | 163.28 | 662.83 | 7.69 | 0.005 |
| 2hab + ssn | 3hab × rnd | 8 | 163.34 | 662.88 | 7.74 | 0.005 |
| 2hab × ssn | 3hab × rnd | 9 | 162.08 | 663.67 | 8.53 | 0.003 |
| ssn | 3hab | 5 | 170.61 | 664.05 | 8.92 | 0.003 |
| ssn | 3hab + rnd | 6 | 168.87 | 664.34 | 9.20 | 0.002 |
| ssn | 3hab × rnd | 7 | 166.93 | 664.43 | 9.29 | 0.002 |
| 3hab × ssn | 2hab | 8 | 165.02 | 664.56 | 9.42 | 0.002 |
| 3hab | 3hab + rnd | 7 | 168.52 | 666.03 | 10.89 | 0.001 |
| 3hab | 3hab × rnd | 8 | 166.59 | 666.13 | 11.00 | 0.001 |
| 3hab | 3hab | 6 | 171.45 | 666.93 | 11.79 | 0.001 |
| 3hab + ssn | 2hab | 6 | 172.79 | 668.27 | 13.13 | 0.000 |
| 3hab × ssn | con | 7 | 171.17 | 668.67 | 13.54 | 0.000 |
| 2hab + ssn | 2hab | 5 | 177.33 | 670.77 | 15.63 | 0.000 |
| 2hab × ssn | 2hab | 6 | 175.35 | 670.82 | 15.68 | 0.000 |
| 2hab | 3hab × rnd | 7 | 176.13 | 673.64 | 18.50 | 0.000 |
| 2hab | 3hab + rnd | 6 | 178.20 | 673.68 | 18.54 | 0.000 |
| ssn | 2hab | 4 | 182.34 | 673.76 | 18.62 | 0.000 |
| 3hab + ssn | con | 5 | 180.74 | 674.19 | 19.05 | 0.000 |
| 2hab | 3hab | 5 | 180.91 | 674.35 | 19.22 | 0.000 |
| con | 3hab × rnd | 6 | 179.15 | 674.62 | 19.48 | 0.000 |
| 2hab × ssn | con | 5 | 181.38 | 674.83 | 19.69 | 0.000 |
| con | 3hab + rnd | 5 | 181.54 | 674.98 | 19.85 | 0.000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| con | 3hab | 4 | 184.75 | 676.17 | 21.04 | 0.000 |
| 2hab + ssn | con | 4 | 184.95 | 676.36 | 21.23 | 0.000 |
| 3hab | 2hab | 5 | 184.59 | 678.03 | 22.89 | 0.000 |
| 2hab | 2hab | 4 | 190.53 | 681.95 | 26.82 | 0.000 |
| ssn | con | 3 | 193.42 | 682.82 | 27.69 | 0.000 |
| con | 2hab | 3 | 194.64 | 684.04 | 28.90 | 0.000 |
| 3hab | con | 4 | 196.99 | 688.41 | 33.28 | 0.000 |
| 2hab | con | 3 | 202.72 | 692.12 | 36.99 | 0.000 |
| con | con | 2 | 213.23 | 700.62 | 45.48 | 0.000 |
| ĉ = 3 | | | | | | |
| 3hab + ssn | 3hab | 7 | 105.28 | 441.66 | 0.00 | 0.253 |
| 3hab + ssn | 3hab + rnd | 8 | 104.22 | 442.64 | 0.98 | 0.155 |
| 3hab × ssn | 3hab | 9 | 102.37 | 442.83 | 1.17 | 0.141 |
| 3hab + ssn | 3hab × rnd | 9 | 103.32 | 443.78 | 2.12 | 0.088 |
| 3hab × ssn | 3hab + rnd | 10 | 101.55 | 444.06 | 2.40 | 0.076 |
| 2hab + ssn | 3hab | 6 | 110.76 | 445.11 | 3.44 | 0.045 |
| 3hab × ssn | 3hab × rnd | 11 | 100.80 | 445.36 | 3.70 | 0.040 |
| 2hab × ssn | 3hab | 7 | 109.60 | 445.99 | 4.32 | 0.029 |
| ssn | 3hab | 5 | 113.74 | 446.06 | 4.40 | 0.028 |
| 2hab + ssn | 3hab + rnd | 7 | 109.88 | 446.26 | 4.59 | 0.025 |
| ssn | 3hab + rnd | 6 | 112.58 | 446.93 | 5.26 | 0.018 |
| 2hab × ssn | 3hab + rnd | 8 | 108.86 | 447.28 | 5.61 | 0.015 |
| 2hab + ssn | 3hab × rnd | 8 | 108.89 | 447.31 | 5.65 | 0.015 |
| ssn | 3hab × rnd | 7 | 111.28 | 447.67 | 6.00 | 0.013 |
| 3hab × ssn | 2hab | 8 | 110.01 | 448.43 | 6.77 | 0.009 |
| 2hab × ssn | 3hab × rnd | 9 | 108.06 | 448.52 | 6.86 | 0.008 |
| 3hab | 3hab | 6 | 114.30 | 448.65 | 6.99 | 0.008 |
| 3hab | 3hab + rnd | 7 | 112.35 | 448.73 | 7.07 | 0.007 |
| 3hab | 3hab × rnd | 8 | 111.06 | 449.48 | 7.82 | 0.005 |
| 3hab + ssn | 2hab | 6 | 115.20 | 449.54 | 7.88 | 0.005 |
| 3hab × ssn | con | 7 | 114.11 | 450.49 | 8.83 | 0.003 |
| 2hab + ssn | 2hab | 5 | 118.22 | 450.54 | 8.88 | 0.003 |
| 2hab × ssn | 2hab | 6 | 116.90 | 451.25 | 9.58 | 0.002 |
| ssn | 2hab | 4 | 121.56 | 451.85 | 10.19 | 0.002 |
| 3hab + ssn | con | 5 | 120.50 | 452.82 | 11.15 | 0.001 |
| 2hab | 3hab | 5 | 120.61 | 452.93 | 11.26 | 0.001 |
| 2hab | 3hab + rnd | 6 | 118.80 | 453.15 | 11.49 | 0.001 |
| 2hab × ssn | con | 5 | 120.92 | 453.24 | 11.58 | 0.001 |
| con | 3hab + rnd | 5 | 121.03 | 453.35 | 11.68 | 0.001 |
| con | 3hab | 4 | 123.17 | 453.46 | 11.80 | 0.001 |
| 2hab + ssn | con | 4 | 123.30 | 453.59 | 11.93 | 0.001 |
| con | 3hab × rnd | 6 | 119.43 | 453.78 | 12.12 | 0.001 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2hab | 3hab × rnd | 7 | 117.42 | 453.80 | 12.14 | 0.001 |
| 3hab | 2hab | 5 | 123.06 | 455.38 | 13.72 | 0.000 |
| ssn | con | 3 | 128.95 | 457.23 | 15.56 | 0.000 |
| 2hab | 2hab | 4 | 127.02 | 457.32 | 15.66 | 0.000 |
| con | 2hab | 3 | 129.76 | 458.04 | 16.37 | 0.000 |
| 3hab | con | 4 | 131.33 | 461.62 | 19.96 | 0.000 |
| 2hab | con | 3 | 135.15 | 463.43 | 21.76 | 0.000 |
| con | con | 2 | 142.15 | 468.42 | 26.75 | 0.000 |
| ĉ = 4 | | | | | | |
| 3hab + ssn | 3hab | 7 | 78.96 | 334.78 | 0.00 | 0.210 |
| 3hab + ssn | 3hab + rnd | 8 | 78.17 | 336.03 | 1.24 | 0.113 |
| 3hab × ssn | 3hab | 9 | 76.78 | 336.68 | 1.90 | 0.081 |
| 2hab + ssn | 3hab | 6 | 83.07 | 336.85 | 2.07 | 0.074 |
| ssn | 3hab | 5 | 85.30 | 337.06 | 2.28 | 0.067 |
| 3hab + ssn | 3hab × rnd | 9 | 77.49 | 337.39 | 2.61 | 0.057 |
| 2hab × ssn | 3hab | 7 | 82.20 | 338.02 | 3.24 | 0.041 |
| 3hab × ssn | 3hab + rnd | 10 | 76.17 | 338.11 | 3.33 | 0.040 |
| ssn | 3hab + rnd | 6 | 84.43 | 338.22 | 3.44 | 0.038 |
| 2hab + ssn | 3hab + rnd | 7 | 82.41 | 338.23 | 3.45 | 0.037 |
| ssn | 3hab × rnd | 7 | 83.46 | 339.28 | 4.50 | 0.022 |
| 2hab × ssn | 3hab + rnd | 8 | 81.64 | 339.50 | 4.72 | 0.020 |
| 3hab | 3hab | 6 | 85.73 | 339.51 | 4.73 | 0.020 |
| 2hab + ssn | 3hab × rnd | 8 | 81.67 | 339.53 | 4.75 | 0.020 |
| 3hab × ssn | 3hab × rnd | 11 | 75.60 | 339.60 | 4.82 | 0.019 |
| 3hab | 3hab + rnd | 7 | 84.26 | 340.08 | 5.30 | 0.015 |
| 3hab + ssn | 2hab | 6 | 86.40 | 340.18 | 5.40 | 0.014 |
| 3hab × ssn | 2hab | 8 | 82.51 | 340.37 | 5.59 | 0.013 |
| 2hab + ssn | 2hab | 5 | 88.66 | 340.42 | 5.64 | 0.012 |
| ssn | 2hab | 4 | 91.17 | 340.90 | 6.12 | 0.010 |
| 2hab × ssn | 3hab × rnd | 9 | 81.04 | 340.94 | 6.16 | 0.010 |
| 3hab | 3hab × rnd | 8 | 83.29 | 341.15 | 6.37 | 0.009 |
| 3hab × ssn | con | 7 | 85.58 | 341.40 | 6.62 | 0.008 |
| 2hab × ssn | 2hab | 6 | 87.67 | 341.46 | 6.68 | 0.007 |
| con | 3hab | 4 | 92.38 | 342.11 | 7.33 | 0.005 |
| 3hab + ssn | con | 5 | 90.37 | 342.13 | 7.35 | 0.005 |
| 2hab + ssn | con | 4 | 92.47 | 342.21 | 7.43 | 0.005 |
| 2hab | 3hab | 5 | 90.45 | 342.21 | 7.43 | 0.005 |
| 2hab × ssn | con | 5 | 90.69 | 342.45 | 7.67 | 0.005 |
| con | 3hab + rnd | 5 | 90.77 | 342.53 | 7.75 | 0.004 |
| 2hab | 3hab + rnd | 6 | 89.10 | 342.89 | 8.11 | 0.004 |
| con | 3hab × rnd | 6 | 89.57 | 343.36 | 8.58 | 0.003 |
| 2hab | 3hab × rnd | 7 | 88.07 | 343.88 | 9.10 | 0.002 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3hab | 2hab | 5 | 92.29 | 344.05 | 9.27 | 0.002 |
| ssn | con | 3 | 96.71 | 344.43 | 9.65 | 0.002 |
| 2hab | 2hab | 4 | 95.27 | 345.00 | 10.22 | 0.001 |
| con | 2hab | 3 | 97.32 | 345.03 | 10.25 | 0.001 |
| 3hab | con | 4 | 98.50 | 348.23 | 13.45 | 0.000 |
| 2hab | con | 3 | 101.36 | 349.08 | 14.30 | 0.000 |
| con | con | 2 | 106.62 | 352.32 | 17.54 | 0.000 |
| ĉ = 5 | | | | | | |
| 3hab + ssn | 3hab | 7 | 63.17 | 270.65 | 0.00 | 0.153 |
| ssn | 3hab | 5 | 68.24 | 271.66 | 1.01 | 0.092 |
| 2hab + ssn | 3hab | 6 | 66.45 | 271.90 | 1.25 | 0.082 |
| 3hab + ssn | 3hab + rnd | 8 | 62.53 | 272.05 | 1.40 | 0.076 |
| 3hab × ssn | 3hab | 9 | 61.42 | 272.98 | 2.33 | 0.048 |
| ssn | 3hab + rnd | 6 | 67.55 | 273.00 | 2.34 | 0.047 |
| 2hab × ssn | 3hab | 7 | 65.76 | 273.24 | 2.59 | 0.042 |
| 2hab + ssn | 3hab + rnd | 7 | 65.93 | 273.41 | 2.76 | 0.039 |
| 3hab + ssn | 3hab × rnd | 9 | 61.99 | 273.56 | 2.91 | 0.036 |
| 3hab | 3hab | 6 | 68.58 | 274.03 | 3.38 | 0.028 |
| ssn | 3hab × rnd | 7 | 66.77 | 274.25 | 3.60 | 0.025 |
| ssn | 2hab | 4 | 72.94 | 274.33 | 3.68 | 0.024 |
| 2hab + ssn | 2hab | 5 | 70.93 | 274.35 | 3.70 | 0.024 |
| 3hab × ssn | 3hab + rnd | 10 | 60.93 | 274.54 | 3.89 | 0.022 |
| 3hab + ssn | 2hab | 6 | 69.12 | 274.57 | 3.92 | 0.022 |
| 2hab × ssn | 3hab + rnd | 8 | 65.31 | 274.83 | 4.18 | 0.019 |
| 2hab + ssn | 3hab × rnd | 8 | 65.34 | 274.86 | 4.20 | 0.019 |
| 3hab | 3hab + rnd | 7 | 67.41 | 274.89 | 4.24 | 0.018 |
| con | 3hab | 4 | 73.90 | 275.30 | 4.65 | 0.015 |
| 2hab + ssn | con | 4 | 73.98 | 275.37 | 4.72 | 0.014 |
| 3hab × ssn | 2hab | 8 | 66.01 | 275.53 | 4.88 | 0.013 |
| 2hab × ssn | 2hab | 6 | 70.14 | 275.59 | 4.94 | 0.013 |
| 3hab + ssn | con | 5 | 72.30 | 275.72 | 5.07 | 0.012 |
| 2hab | 3hab | 5 | 72.36 | 275.78 | 5.13 | 0.012 |
| 3hab × ssn | con | 7 | 68.47 | 275.95 | 5.30 | 0.011 |
| 2hab × ssn | con | 5 | 72.55 | 275.97 | 5.32 | 0.011 |
| con | 3hab + rnd | 5 | 72.62 | 276.04 | 5.39 | 0.010 |
| 3hab × ssn | 3hab × rnd | 11 | 60.48 | 276.14 | 5.49 | 0.010 |
| 3hab | 3hab × rnd | 8 | 66.64 | 276.16 | 5.51 | 0.010 |
| 2hab × ssn | 3hab × rnd | 9 | 64.83 | 276.40 | 5.75 | 0.009 |
| 2hab | 3hab + rnd | 6 | 71.28 | 276.73 | 6.08 | 0.007 |
| ssn | con | 3 | 77.37 | 276.75 | 6.10 | 0.007 |
| con | 3hab × rnd | 6 | 71.66 | 277.11 | 6.46 | 0.006 |
| con | 2hab | 3 | 77.86 | 277.23 | 6.58 | 0.006 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3hab | 2hab | 5 | 73.84 | 277.26 | 6.60 | 0.006 |
| 2hab | 2hab | 4 | 76.21 | 277.61 | 6.96 | 0.005 |
| 2hab | 3hab × rnd | 7 | 70.45 | 277.93 | 7.28 | 0.004 |
| 3hab | con | 4 | 78.80 | 280.19 | 9.54 | 0.001 |
| 2hab | con | 3 | 81.09 | 280.47 | 9.82 | 0.001 |
| con | con | 2 | 85.29 | 282.66 | 12.01 | 0.000 |
| ĉ = 6 | | | | | | |
| 3hab + ssn | 3hab | 7 | 52.64 | 227.90 | 0.00 | 0.110 |
| ssn | 3hab | 5 | 56.87 | 228.06 | 0.17 | 0.101 |
| 2hab + ssn | 3hab | 6 | 55.38 | 228.60 | 0.70 | 0.077 |
| 3hab + ssn | 3hab + rnd | 8 | 52.11 | 229.41 | 1.51 | 0.052 |
| ssn | 3hab + rnd | 6 | 56.29 | 229.51 | 1.62 | 0.049 |
| ssn | 2hab | 4 | 60.78 | 229.95 | 2.05 | 0.039 |
| 2hab × ssn | 3hab | 7 | 54.80 | 230.06 | 2.16 | 0.037 |
| 2hab + ssn | 3hab + rnd | 7 | 54.94 | 230.19 | 2.30 | 0.035 |
| 2hab + ssn | 2hab | 5 | 59.11 | 230.30 | 2.41 | 0.033 |
| 3hab | 3hab | 6 | 57.15 | 230.38 | 2.48 | 0.032 |
| 3hab × ssn | 3hab | 9 | 51.18 | 230.52 | 2.62 | 0.030 |
| con | 3hab | 4 | 61.58 | 230.76 | 2.86 | 0.026 |
| 2hab + ssn | con | 4 | 61.65 | 230.82 | 2.92 | 0.025 |
| 3hab + ssn | 2hab | 6 | 57.60 | 230.82 | 2.92 | 0.025 |
| ssn | 3hab × rnd | 7 | 55.64 | 230.90 | 3.00 | 0.024 |
| 3hab + ssn | 3hab × rnd | 9 | 51.66 | 231.00 | 3.10 | 0.023 |
| 3hab | 3hab + rnd | 7 | 56.17 | 231.43 | 3.53 | 0.019 |
| 3hab + ssn | con | 5 | 60.25 | 231.44 | 3.55 | 0.019 |
| 2hab | 3hab | 5 | 60.30 | 231.50 | 3.60 | 0.018 |
| ssn | con | 3 | 64.47 | 231.63 | 3.73 | 0.017 |
| 2hab × ssn | con | 5 | 60.46 | 231.66 | 3.76 | 0.017 |
| 2hab × ssn | 2hab | 6 | 58.45 | 231.67 | 3.78 | 0.017 |
| con | 3hab + rnd | 5 | 60.51 | 231.71 | 3.81 | 0.016 |
| 2hab × ssn | 3hab + rnd | 8 | 54.43 | 231.72 | 3.83 | 0.016 |
| 2hab + ssn | 3hab × rnd | 8 | 54.45 | 231.74 | 3.84 | 0.016 |
| con | 2hab | 3 | 64.88 | 232.03 | 4.13 | 0.014 |
| 3hab × ssn | 3hab + rnd | 10 | 50.78 | 232.16 | 4.27 | 0.013 |
| 3hab × ssn | 2hab | 8 | 55.01 | 232.30 | 4.40 | 0.012 |
| 3hab × ssn | con | 7 | 57.06 | 232.31 | 4.41 | 0.012 |
| 2hab | 3hab + rnd | 6 | 59.40 | 232.62 | 4.73 | 0.010 |
| 2hab | 2hab | 4 | 63.51 | 232.68 | 4.78 | 0.010 |
| 3hab | 2hab | 5 | 61.53 | 232.72 | 4.83 | 0.010 |
| 3hab | 3hab × rnd | 8 | 55.53 | 232.82 | 4.93 | 0.009 |
| con | 3hab × rnd | 6 | 59.72 | 232.94 | 5.04 | 0.009 |
| 2hab × ssn | 3hab × rnd | 9 | 54.03 | 233.37 | 5.47 | 0.007 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 3hab × ssn | 3hab × rnd | 11 | 50.40 | 233.84 | 5.94 | 0.006 |
| 2hab | 3hab × rnd | 7 | 58.71 | 233.97 | 6.07 | 0.005 |
| 2hab | con | 3 | 67.57 | 234.73 | 6.83 | 0.004 |
| 3hab | con | 4 | 65.66 | 234.84 | 6.94 | 0.003 |
| con | con | 2 | 71.08 | 236.22 | 8.32 | 0.002 |

F. Assessment of the reliability of song similarity scoring methods

Table S6. Differences between within individual and across individual measures of song similarity (± 95% bootstrap confidence interval) by site. All differences are > 0 demonstrating that, on average, within-individual song comparisons generated higher song similarity values than among-individual song comparisons within the same site.

| site | lower confidence limit | | average difference | | upper confidence limit |
|---|---|---|---|---|---|
| grassland: | | | | | |
| 2A-R20A | 0.077 | – | 0.109 | – | 0.140 |
| 4F | 0.069 | – | 0.101 | – | 0.131 |
| 20C | 0.009 | – | 0.042 | – | 0.074 |
| C1B | 0.067 | – | 0.097 | – | 0.127 |
| cropland: | | | | | |
| Ashland | 0.055 | – | 0.105 | – | 0.150 |
| North | 0.110 | – | 0.162 | – | 0.214 |
| Strauss | 0.203 | – | 0.238 | – | 0.271 |

Exploratory analysis conducted with cross-correlation

We conducted a preliminary study to evaluate the suitability of the cross-correlation in Raven for quantifying song similarity in dickcissels. In this preliminary analysis, conducted in 2010, we used songs of color-banded territorial male Dickcissels recorded in the summers of 2006-2010 at our sites described in the main text of this paper.  The within-bird portion of our analysis included 24 dickcissels present in multiple years of the study from four sites: two sites on the Konza Prairie Biological Station (KPBS) and two agricultural sites managed by Kansas State University (KSU).  This portion included 54 song comparisons.  The between-bird portion of our analysis included 24 birds from the same sites and 2851 song comparisons.

*Evaluation Methods*
We used RavenPro 1.3 (Cornell Lab of Ornithology 2008) to convert high-quality song recordings into visual spectrograms for analysis.  We used a bandpass filter to eliminate frequencies above 20,000 Hz and below 2,000 Hz.  We removed isolated insect noises and the songs of other birds using the select function where they did not interfere with the target bird's song.

Before evaluating song similarity, we focused on cissel phrases.  For every date we visited a site, we chose the cleanest recording of each cissel phrase of each birds' song to include in cross-correlation analysis.  Even if other phrases were present in the song, we only looked at dick and cissel phrases as they appeared in some variety in every bird's song.

Using Raven, we performed batch cross-correlations of each phrase within each site and between the sites producing a large matrix of cross-correlation values comparing all cissel phrases to all cissel phrases.

*Preliminary Findings: Within-bird v. Between Bird*
The cross-correlation results comparing all birds present for multiple years showed a trend for higher cross-correlation scores within individual birds than between birds; but ultimately within-bird cross-correlation scores should have been consistently higher than between-bird cross-correlation scores. Both groups had substantial unexplained variability in cross-correlation score distribution, with the same range of scores for within-bird comparisons as in between-bird comparisons. Within-individual comparisons produced scores as low as 0.1 and many scores < 0.6 (Figures S1, S2).

Figure S2

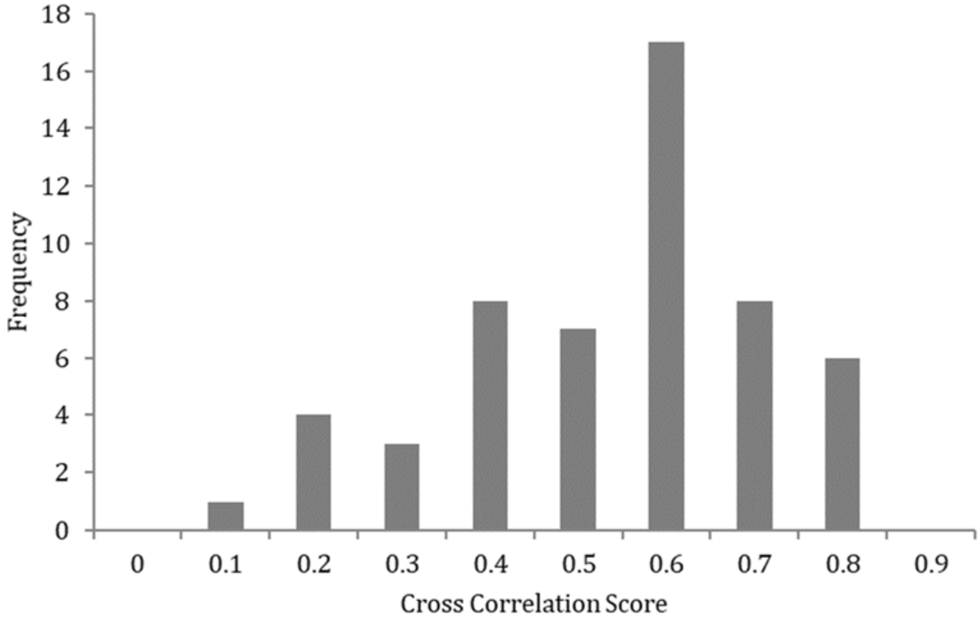


Figure S2. Cross correlation scores for comparisons between two songs produced by the same male dickcissel on different dates.

Figure S3

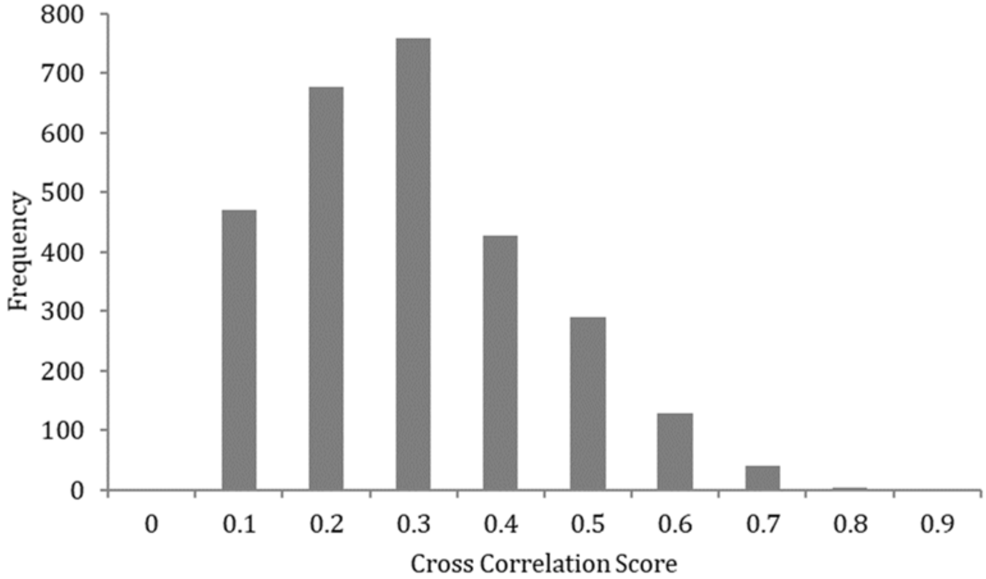


Figure S3. Cross correlation scores for comparisons between pairs of songs sung by different male dickcissel.

G. Model-averaged parameter estimates from mark-recapture models

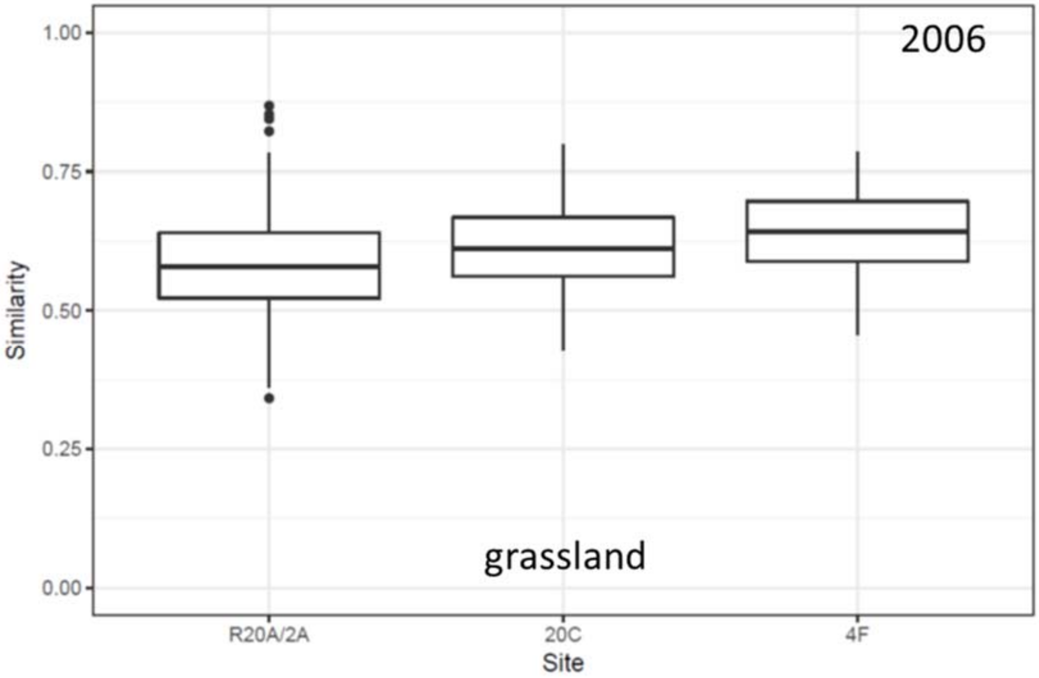
Table S7. Model-averaged parameter estimates associated with all factor levels for variables included in candidate models in mark-recapture analyses of male dickcissels.

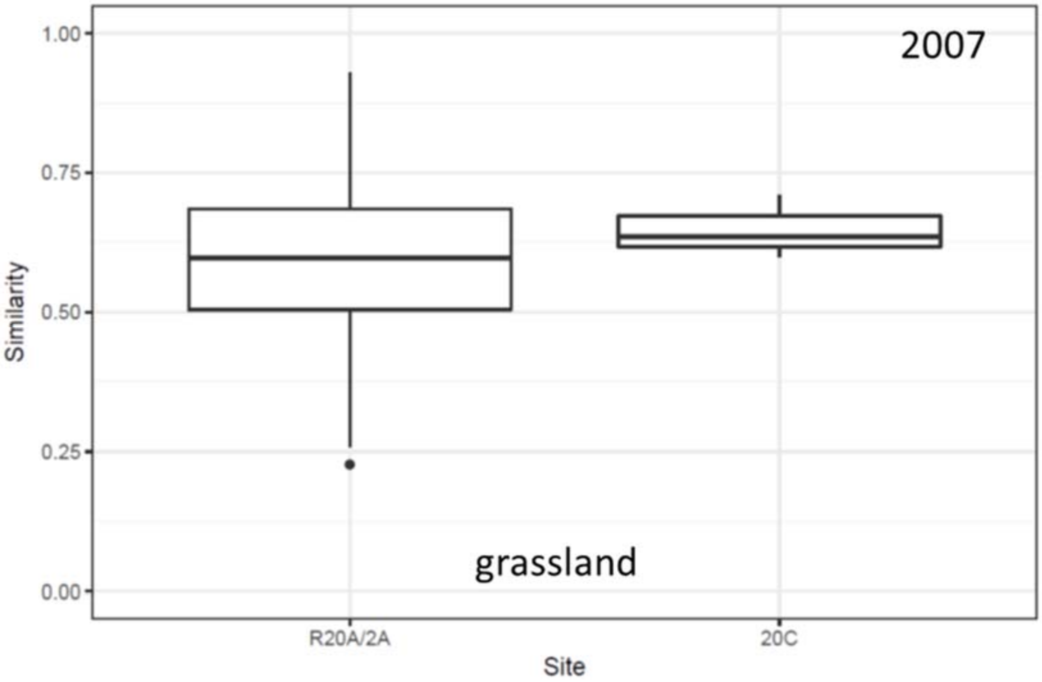
| factor and level | weighted average phi (daily) | SE |
|---|---|---|
| breeding season (3 hab models) | | |
| Cropland | 0.9849295 | 0.0038969 |
| Grassland (T. Parker) | 0.9946631 | 0.0013790 |
| Grassland (B. Sousa) | 0.9934902 | 0.0018217 |
| non-breeding season (3 hab models) | | |
| Cropland | 0.9976892 | 0.0008993 |
| Grassland (T. Parker) | 0.9989548 | 0.0002517 |
| Grassland (B. Sousa) | 0.9967646 | 0.0005383 |
| breeding season (2 hab models) | | |
| Cropland breeding season | 0.9850170 | 0.0039555 |
| Grassland breeding season | 0.9946955 | 0.0010306 |
| non-breeding season (2 hab models) | | |
| Cropland non-breeding season | 0.9975199 | 0.0009614 |
| Grassland non-breeding season | 0.9981919 | 0.0002413 |
| p | weighted average p | SE |
| (3 hab models) | | |
| Cropland, first occasion of the year | 0.7060629 | 0.0811538 |
| Cropland, subsequent occasions of the year | 0.5912028 | 0.0748779 |
| Grassland (T. Parker), first occasions | 0.7850901 | 0.0447466 |
| Grassland (T. Parker), subsequent occasions | 0.7521143 | 0.0289636 |
| Grassland (B. Sousa), first occasions | 0.9283292 | 0.0210338 |
| Grassland (B. Sousa), subsequent occasions | 0.9283292 | 0.0210338 |
| (2 hab models) | | |
| Cropland, all occasions | 0.6378040 | 0.0577879 |

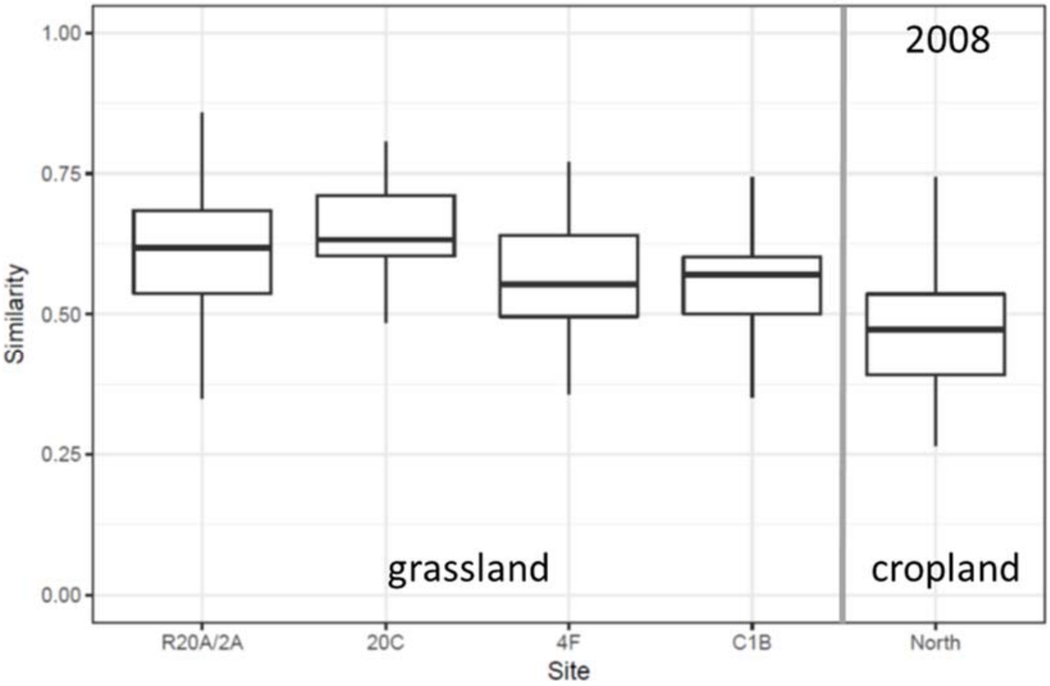
| | | |
|---|---|---|
| Grassland, all occasions | 0.8273261 | 0.0173477 |

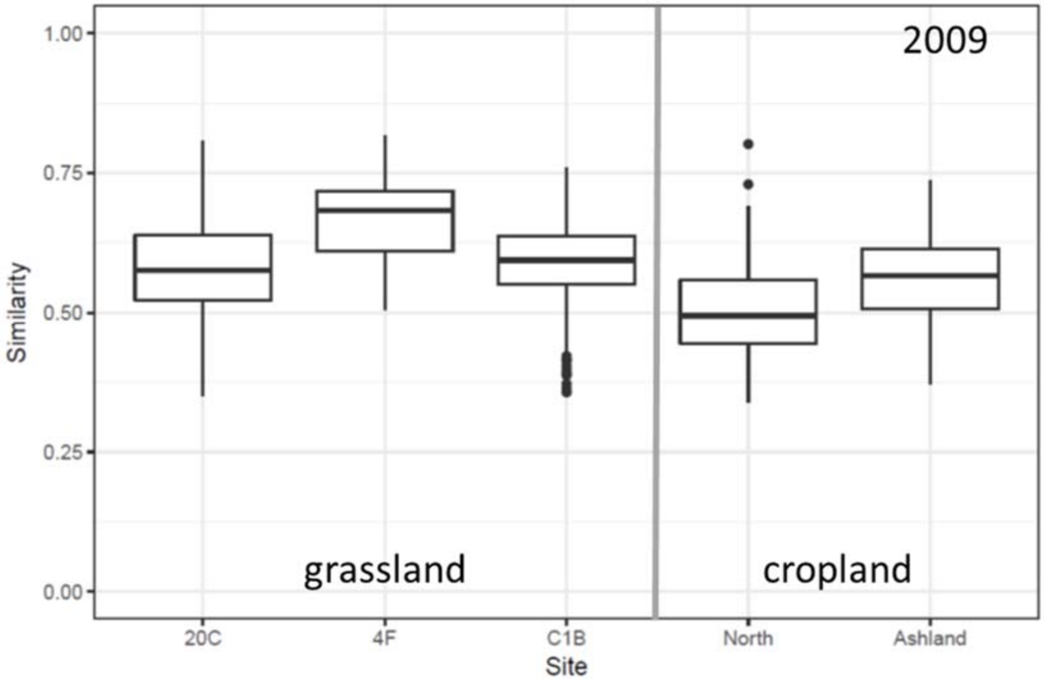
H. Graphical evidence of patterns in vocal culture

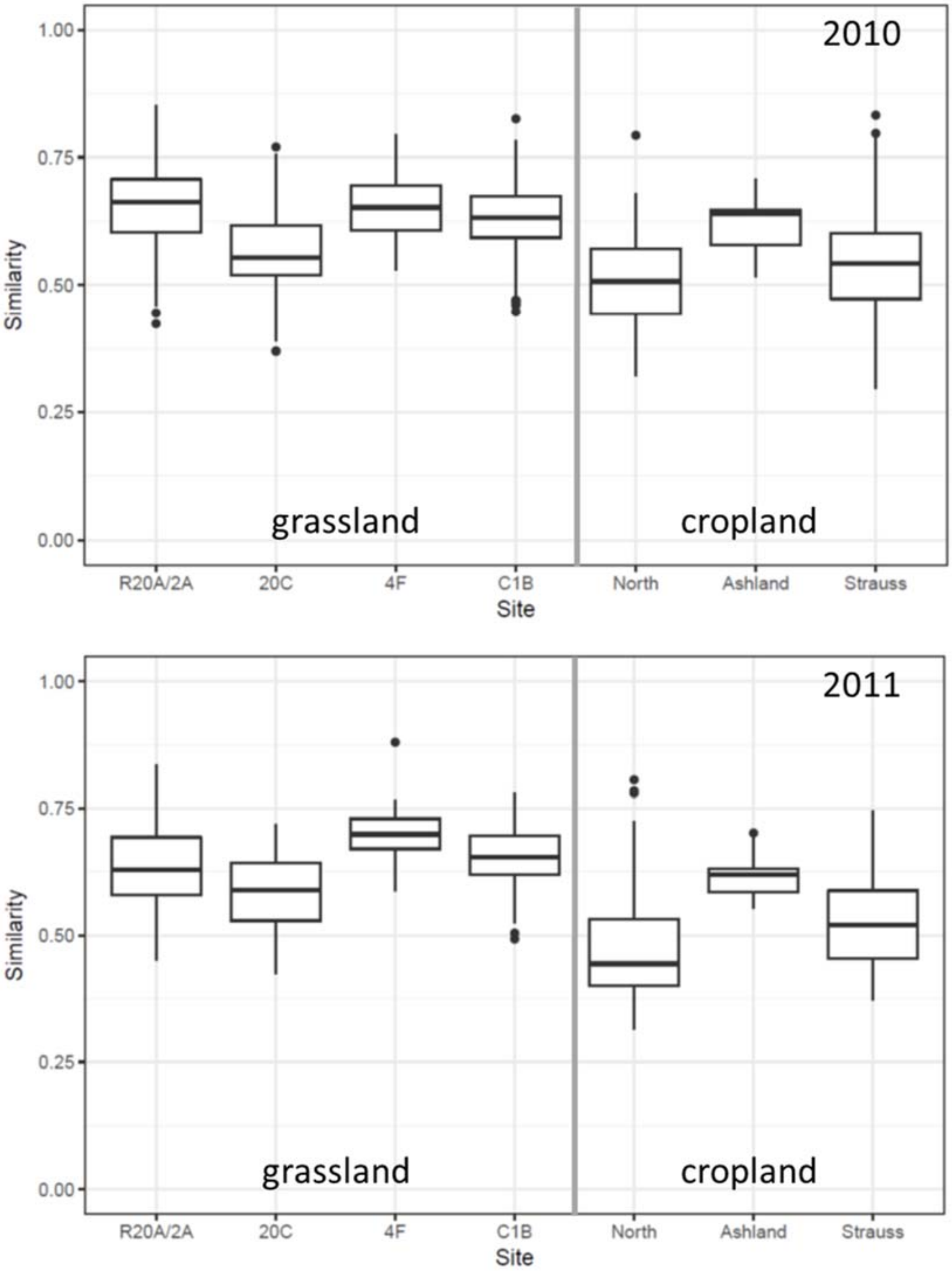
Figure S4. Plots of median song similarity values across the grassland sites (R20A/2A, 20C, 4F, C1B) and cropland sites (North, Ashland, Strauss) in different years. The central line is the median and the box represents the interquartile range (middle 50th percentile).






2008
Similarity
1.00
0.75
0.50
0.25
0.00
grassland
cropland
R20A/2A
20C
4F
C1B
North
Site


2009
Similarity
1.00
0.75
0.50
0.25
0.00
grassland
cropland
20C
4F
C1B
North
Ashland
Site


2010
Similarity
1.00
0.75
0.50
0.25
0.00
grassland
cropland
R20A/2A
20C
4F
C1B
North
Ashland
Strauss
Site
2011
Similarity
1.00
0.75
0.50
0.25
0.00
grassland
cropland
R20A/2A
20C
4F
C1B
North
Ashland
Strauss
Site

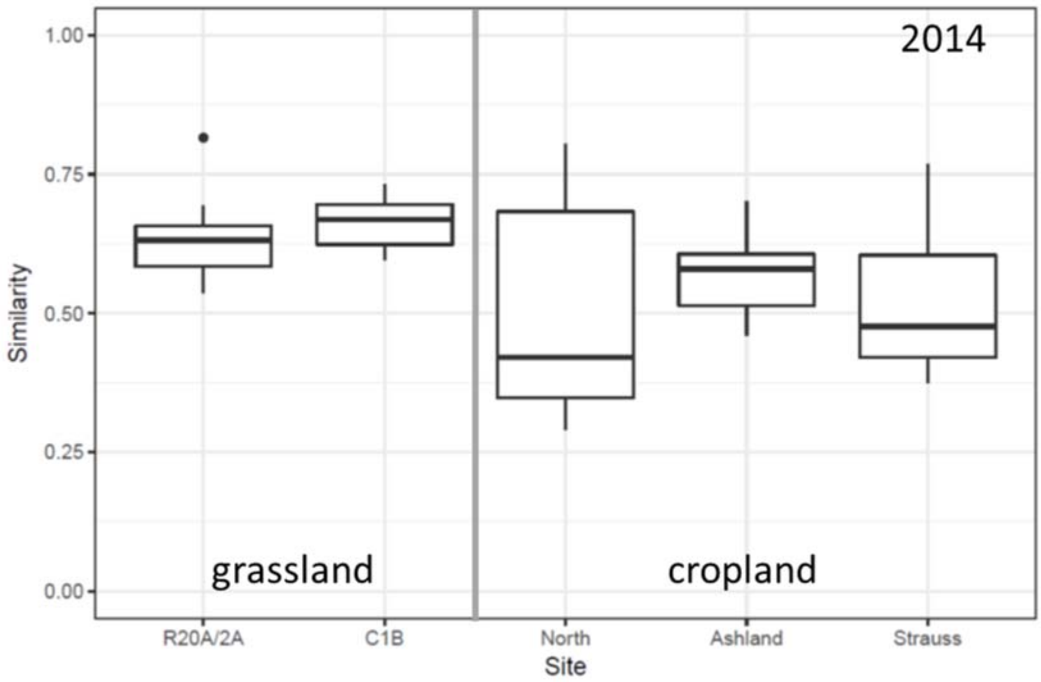
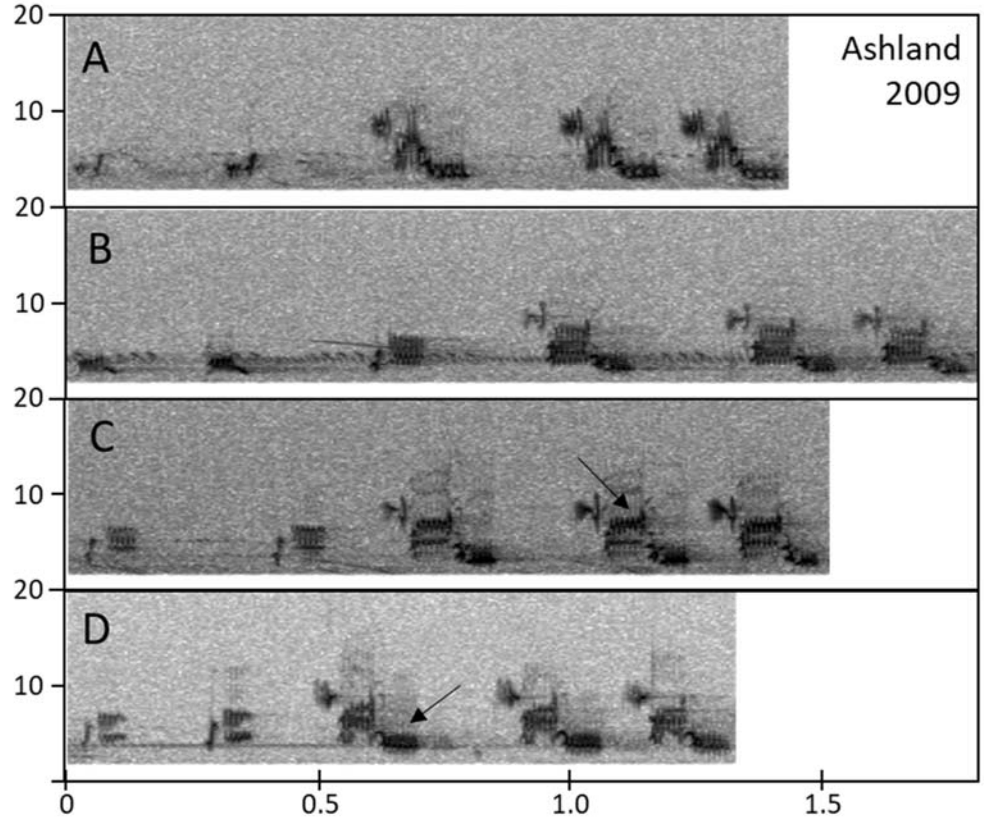
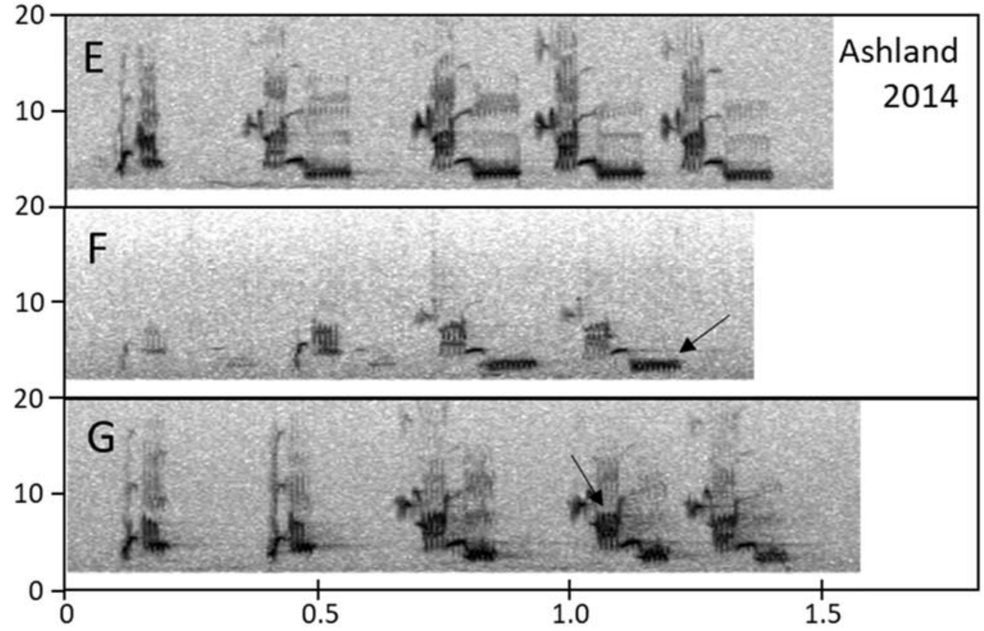
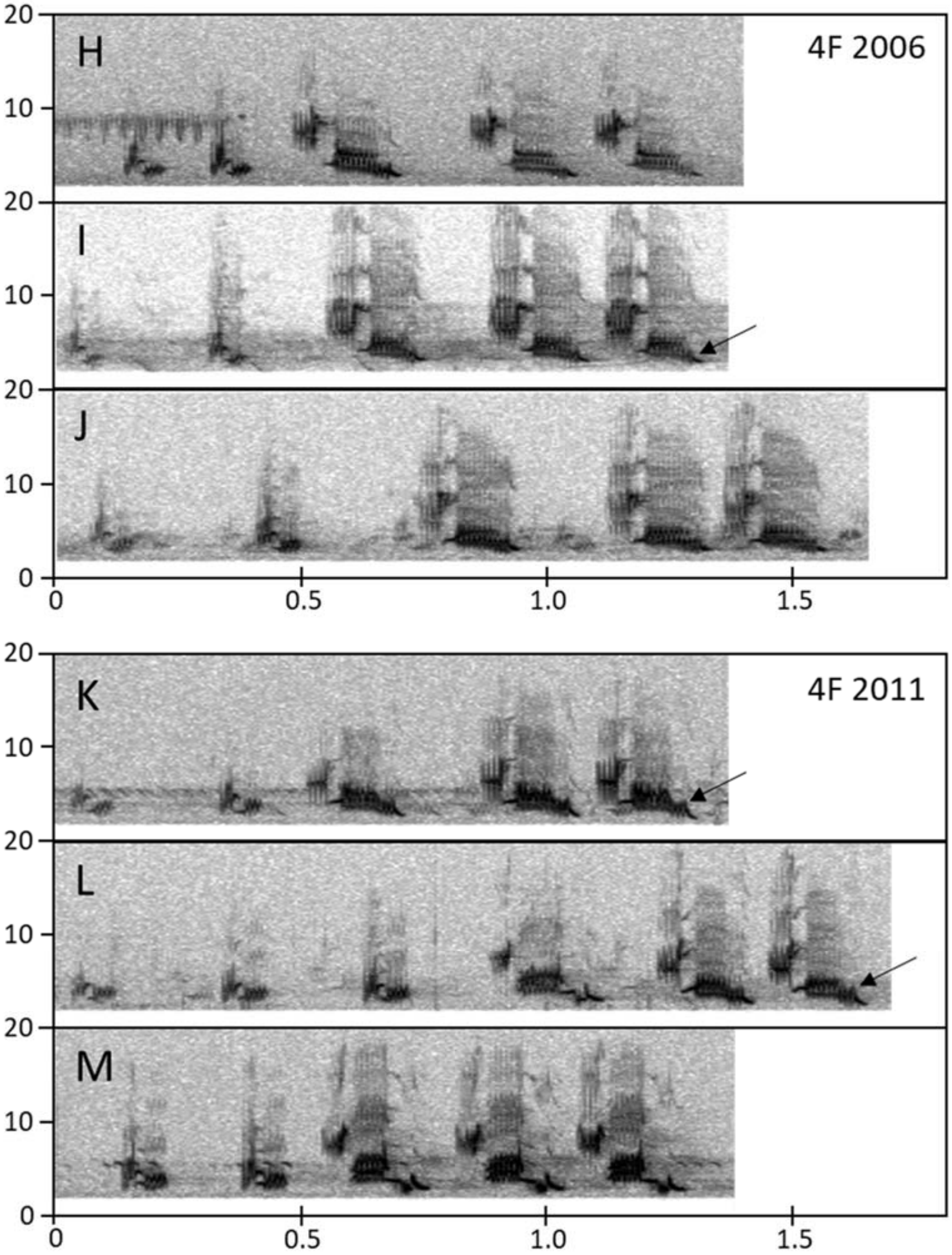

2014
Similarity
1.00
0.75
0.50
0.25
0.00
grassland
cropland
R20A/2A
C1B
North
Ashland
Strauss
Site

Figure S5. Changes in vocal culture of dickcissels over five years at two different sites (Ashland, a cropland site, 2009 and 2014; 4F a grassland site, 2006 and 2011). In most Ashland songs in 2014, the central portion of cissel phrase (arrow in panel G) was of shorter duration than the comparable portion in 2009 (arrow in panel C). Also, the low frequency terminal portion of the cissel phrase tended to be much longer and to have a narrower frequency range in 2014 (arrow in panel F) than in 2009 (arrow in panel D). Some of the evident differences in 4F songs are more subtle. For instance, the final note of the common cissel phrase in 2006 (arrow in panel I) was of slightly longer duration and greater frequency range in 2011 (arrow in panel K). However, a distinctly different cissel phrase can also be seen in some 2011 songs. Some males in 2011 only produced this different type of cissel (panel M), but some other males produced both types of cissels (panel L).


A
Ashland
2009
B
C
D
20
10
0
0.5
1.0
1.5


E
Ashland
2014
F
G
20
10
0
0
0.5
1.0
1.5


H
4F 2006
I
J
K
4F 2011
L
M
20
10
0
0
0.5
1.0
1.5

## I. Literature Cited

Krebs, J. R., and D. E. Kroodsma. 1980. Repertoires and geographic variation in bird song. Pages 143-177 in J. S. Rosenblatt, R. A. Hinde, C. Beer, and M. C. Busnel, editors. Advances in the Study of Behavior, 11. Academic Press, New York.

Parra, M., A. C. Dalisio, W. E. Jensen, and T. H. Parker. 2017. Male territorial aggression does not drive conformity to local vocal culture in a passerine bird. Ethology 123:800-810.

Schook, D. M., M. D. Collins, W. E. Jensen, P. J. Williams, N. E. Bader, and T. H. Parker. 2008. Geographic patterns of song similarity in the dickcissel, Spiza americana. Auk 125:953-964.

Small, D. M., M. E. Gimpel, and D. E. Gill. 2012. Site fidelity and natal philopatry in dickcissels. Northeastern Naturalist 19:123-129.

Sousa, B. F., and D. F. Westneat. 2013a. Positive association between social and extra-pair mating in a polygynous songbird, the dickcissel (Spiza americana). Behavioral Ecology and Sociobiology 67:243-255.

Sousa, B. F., and D. F. Westneat. 2013b. Variance in mating success does not produce strong sexual selection in a polygynous songbird. Behavioral Ecology 24:1381-1389.

Temple, S. A. 2002. Dickcissel (Spiza americana).in A. Poole and F. Gill, editors. The Birds of North America, No. 703. The Birds of North America, Inc., Philadelphia.

Verheijen, B. H. F. 2017. Demographic responses of grassland songbirds to rangeland management in the tallgrass prairie. Kansas State University, Manhattan, Kansas, USA.