ON THE MAINTENANCE OF BIRD SONG DIALECTS

ROBERT PLANQUÉ, NICHOLAS F. BRITTON, AND HANS SLABBEKOORN

keywords: mathematical biology, dialects, bird song, modelling, difference equations, behavioural ecology

Abstract. Many bird species, especially song birds but also for instance some hummingbirds and parrots, have noted dialects. By this we mean that locally a particular song is sung by the majority of the birds, but that neighbouring patches may feature different song types. Behavioural ecologists have been interested in how such dialects come about and how they are maintained for over 45 years. As a result, a great deal is known about different mechanisms at play, such as dispersal, assortative mating and learning of songs, and there are several competing hypotheses to explain the dialect patterns known in nature. There is, however, surprisingly little theoretical work testing these different hypotheses at present. We analyse the simplest kind of model that takes into account the most important biological mechanisms, and in which one may speak of dialects: a model in which there are but two patches, and two song types. It teaches us that a combination of little dispersal and strong assortative mating ensures dialects are maintained. Assuming a simple, linear frequency-dependent learning rule has little effect on the maintenance of dialects. A nonlinear learning rule, however, has dramatic consequences and greatly facilitates dialect maintenance. Adding fitness benefits for singing particular songs in a given patch also has a great impact. Now rare song types may invade and remain in the population.

1. Introduction

Acoustic signals play an important role in structuring animal populations. Both male-male competition and female mate choice are often strongly mediated by acoustic communication. Birds are particularly well known for their advertisement songs and the role they play in sexual selection (Catchpole & Slater, 1995; Marler & Slabbekoorn, 2004). Bird songs vary both within and between species. Intraspecific variation across different geographic localities occurs in practically all avian taxa that have been studied so far (Mundinger, 1982; Podos & Warren, 2007). Interspecific variation often plays a critical role as a pre-zygotic barrier to hybridization between closely related taxa. Under specific conditions, geographic variation within species can be regarded as a precursor of reproductively isolating variation between species, as divergent signals do not have to but can promote speciation through an impact on male settlement success and assortative mating (Slabbekoorn & Smith, 2002; Edwards et al., 2005; Price, 2008, p. 470).

The nature of geographic variation in bird songs varies markedly among taxa. Non-learning species, such as quails, doves, or owls may exhibit gradual, clinal, variation in spectral or temporal parameters (e.g., Goldstein 1978; Baptista et al. 1983; Appleby & Redpath 1997). However, most attention has been drawn by geographic variation in learning species such as songbirds (e.g., Borror (1956); Marler & Tamura

(1962); Payne (1978); Kroodmsa (1985); Handford (1988)), and more recently also hummingbirds (Gaunt et al., 1994; Yang et al., 2007), and parrots (Wright, 1996; Baker, 2003). Songbirds have been best studied, and there is huge variation within this familiar group of bird species in how songs vary geographically (Mundinger, 1982; Podos & Warren, 2007). Detailed studies on species with relatively small repertoire size have shown isolation-by-distance patterns of geographic variation in sharing of song types and similarity in structural song characteristics, with the highest similarity among direct neighbours (Rivera-Gutierrez et al., 2010) or among individuals that are a few territories apart (Lachlan & Slater, 2003). Individuals from other species may have large improvised song repertoires without much geographic structure (Kroodsma et al., 1999) or may typically sing a single stereotypic song, shared among neighbours, and forming small, geographically distinct, acoustic clusters (Slabbekoorn et al., 2003).

Geographically distinct acoustic clusters yield sharp transitions in vocal parameters across localities and are referred to here as dialects (Marler & Tamura, 1962; Baker & Cunningham, 1985; Slabbekoorn & Smith, 2002; Podos & Warren, 2007). The best studied dialectal bird species is the white-crowned sparrow (*Zonotrichia leucophrys*) for which there is ample field data on different subspecies (e.g., Baptista & King 1980; Baker & Cunningham 1985; Cunningham et al. 1989; Nelson et al. 2004). Studies on this species have provided insight into dialect persistence (e.g., Harbison et al. 1999; Nelson et al. 2004), as well as dialect-dependent response behaviour from playbacks in the field (e.g., Nelson et al. 2004; Derryberry 2011), and dialect-dependent local adaptation (MacDougall-Shackleton et al., 2002; Luther & Baptista, 2010). Furthermore, through laboratory studies we know more about song development (e.g., Marler 1970; Nelson 2000), for which we have additional insights from field observations on the impact of social interaction (Baptista & Petrinovich, 1984; Bell et al., 1998; Nelson et al., 2001).

There are also several theoretical studies that have dealt with explaining aspects of dialectal variation not specifically related to a particular species. Such modeling explorations have for example addressed the impact of repertoire size and tutor number on dialect formation (Williams & Slater, 1990); the evolutionary maintenance of song-learning tendency (Lachlan & Slater, 1999); the impact of dialectal variation on geneflow and speciation (Ellers & Slabbekoorn, 2003; Lachlan & Servedio, 2004), and aspects of cultural transmission for multi-component signals (Strigul, 2009). Most of these studies thus address causes and consequences of dialect formation and leave aside the issue of dialect maintenance. However, in the best studied example of white-crowned sparrow, dialects have been shown to persist over a period of forty years (Marler & Tamura, 1962; Nelson et al., 2004). We believe it to be a striking property that dialects are maintained over such long periods of time, given that several factors mentioned above lead to mixing, merging, and dilution of dialect boundaries.

In this theoretical paper, we develop mathematical models to study dialect maintenance through exploring persistence of two adjacent and acoustically distinct populations. The main questions of our paper are 'How are dialect boundaries maintained?' and 'Which factors play a critical role in maintenance?' We shall develop four models, each incorporating the primary factors affecting dialectal dynamics (song-learning, individual dispersal, assortative mating, and environmental selection) in a particular way. The differences between the models are as follows.

Model 1 assumes that birds first disperse to a new territory and then locally learn new song characteristics which replace natal song characteristics to a variable extent (post-dispersal learning). We assume there is no link between song type and the environment. Model 2 assumes that birds disperse after having learnt their song (pre-dispersal learning). Again, there is no link between song type and environment. Model 3 is a variation on Model 1, but now we assume that a particular song type induces fitness benefits associated with certain habitats. Model 4 is another variation on Model 1, in which the learning component is made more realistic by integrating a non-linear rule for the post-dispersal song tutor selection. In each of these models, we study under which conditions dialects persist.

2. Modelling preliminaries

We develop the simplest possible model in which geographic variation in song may occur: one with only two sites 1 and 2, and two song types, which we call dialect 1 and 2. Note that for the particular question of dialect *maintenance* it is not necessary to consider changes in song type, which might be due to copying errors in the learning process or innovations of individual young birds. Such variations only dilute any existing dialect pattern. For the formation of dialects such variations are of course crucial. In small populations, stochastic effects will eventually drive one of two static song types to extinction in the absence of some variation. Our models deal with infinite populations and are deterministic, so that such extinction events will not occur.

We now list the main ingredients for each of the models. At any point in time each individual knows either dialect 1 or dialect 2, but not both. Juvenile birds initially learn a natal dialect, but may later learn the other dialect. (The model for this will be described later.) A bird who learns a new dialect is assumed to forget its old one. (It is known from some species that the old dialect is sometimes retained along with the new one, but for mathematical simplicity this is ignored.) The frequency of individuals at site i who know dialect j at any census point is given by P_{ij}. The model is frequency-independent, and we assume sites are at carrying capacity at each stage. The state of the system at any census point is therefore determined by the variables P_{ij}, for $i = 1, 2$ and $j = 1, 2$. The annual cycle consists of three stages,

- reproduction,
- dispersal of juveniles, and
- learning by juveniles.

Juveniles are assumed to become adult in one year. The census (determining the P_{ij}) is taken before reproduction, and therefore counts one-year-old and older adults, who make up the breeding population in that year. Annual adult mortality is taken to be a constant μ, independent of age. The population is assumed to be at carrying capacity, so that those adults who die in a year are replaced in the breeding population by the same number of one-year-old birds.

We now discuss in detail how reproduction takes place. The implementation of dispersal and learning will be deferred until Model 1, since the order of these two processes (and thus their mathematical implementations) differs between the models.

Assortative mating takes place, in such a way that the number of mixed-dialect matings is less than it would be in a panmictic population. Juvenile birds initially learn the dialect of their father, who sings either dialect with equal probability (so that effectively offspring of mixed-dialect matings have equal probability of learning either dialect); this is known as their *natal dialect.* If p_{ij} is the frequency of juveniles in patch i with natal dialect j, then

$$p_{i1} = \frac{P_{i1}^2 + \sigma P_{i1}P_{i2}}{P_{i1}^2 + 2\sigma P_{i1}P_{i2} + P_{i2}^2}, \quad p_{i2} = \frac{P_{i2}^2 + \sigma P_{i1}P_{i2}}{P_{i1}^2 + 2\sigma P_{i1}P_{i2} + P_{i2}^2}. \tag{1}$$

The parameter σ measures how tolerant the population is to mixed-dialect matings; a value $\sigma = 1$ represents a perfectly panmictic population, while a value $\sigma = 0$ represents one where no mixed-dialect matings take place.

The mathematical analysis in each model consists of finding the equilibria and studying their stability properties. A steady state may be denoted as one with dialects if patch 1 contains a majority of birds singing one dialect type, and patch 2 a majority of the other song type. Due to the symmetric assumptions in the models these majorities are often identical between patches, so e.g., a (80%, 20%) split between dialect 1 and 2 in patch 1 usually means that patch 2 has a (20%, 80%) division. (Strictly speaking, this is not true for all steady states, but it seems true for all stable equilibria.) It is also important to note that the stability of such dialect steady states often coincides with *instability* of the fully mixed (50%, 50%) steady state. In other words, persistence of a dialect pattern is usually equivalent to showing that the fully mixed steady state can segregate into a dialect pattern. This is different from studying when a rare new vocal variant can invade, which would be the typical analysis to study *formation* of new dialects.

3. Model 1: Post-dispersal learning

In the post-dispersal model, juveniles (and only juveniles) are assumed to disperse, with some fixed probability, in such a way that after dispersal a fraction ε of the juveniles at each site have arrived from the other site. (Note that ε is the fraction of *successful* colonisers, which is not necessarily the same as the fraction of *attempted* colonisers.) Let p'_{ij} be the frequency of juveniles at site i with natal dialect j after this dispersal process. Then

$$p'_{1j} = (1-\varepsilon)p_{1j} + \varepsilon p_{2j}, \quad p'_{2j} = (1-\varepsilon)p_{2j} + \varepsilon p_{1j}. \tag{2}$$

The juvenile birds may now keep their natal dialects, with probability $1-\lambda$, or learn a new dialect from an adult bird chosen at random, with probability λ. Let p''_{ij} be the frequency of juveniles at site i with dialect j after learning. Then

$$p''_{ij} = (1-\lambda)p'_{ij} + \lambda P_{ij}. \tag{3}$$

This type of learning is linearly frequency-dependent, sometimes referred to as "linear learning". A non-linear learning rule is explored in Model 4.

To complete the model, recall that a fraction μ of adults die each year and are replaced in the breeding population by the same number of one-year-old birds. Let P'_{ij} be the frequency of birds in patch i with dialect j at the subsequent census. Then Model 1 is given by

$$P'_{ij} = (1-\mu)P_{ij} + \mu p''_{ij}. \tag{4}$$

for each $i = 1, 2$, $j = 1, 2$.

3.1. **Analysis.** Note that P_{ij}, p_{ij}, p'_{ij} and p''_{ij} are frequencies (with respect to j), so we need only to consider new variables $Q_i = P_{i1} - P_{i2}$, $q_i = p_{i1} - p_{i2}$, etc. Then

$$Q'_i = (1-\kappa)Q_i + \kappa q'_i = (1-\kappa)Q_i + \kappa((1-\varepsilon)q_i + \varepsilon q_j), \tag{5}$$

where $\kappa = \mu(1-\lambda)$ and $j \neq i$. It is easy to check that

$$q_i = f(Q_i), \tag{6}$$

where

$$f(Q) = \frac{Q}{\frac{1}{2}(1+\sigma) + \frac{1}{2}(1-\sigma)Q^2}, \tag{7}$$

so the system becomes

$$Q'_i = (1-\kappa)Q_i + \kappa((1-\varepsilon)f(Q_i) + \varepsilon f(Q_j)), \tag{8}$$

where $j \neq i$.

3.1.1. *Symmetric and anti-symmetric steady states.* Note that $f(0) = 0$, $f(1) = 1$, $f(-1) = -1$. There are therefore three steady states that are symmetric under the transformation $(Q_1, Q_2) \to (Q_2, Q_1)$, $Q_1 = Q_2 = 1$ and $Q_1 = Q_2 = -1$, corresponding to extinction of dialect 2 and 1 respectively, and $Q_1 = Q_2 = 0$, corresponding to equal numbers of each dialect in each site, $P_{11} = P_{12} = P_{21} = P_{22} = \frac{1}{2}$. The points $(Q^*, -Q^*)$ and $(-Q^*, Q^*)$ are non-trivial anti-symmetric steady states of the system (under the same transformation) if

$$Q^* = (1-2\varepsilon)f(Q^*), \tag{9}$$

i.e. if

$$\frac{1}{2}(1-\sigma)Q^{*2} = 1 - 2\varepsilon - \frac{1}{2}(1+\sigma), \tag{10}$$

or

$$Q^{*2} = \frac{2(1-2\varepsilon) - (1+\sigma)}{1-\sigma}. \tag{11}$$

A non-trivial pair of such points exists in $\mathbb{R}$ if and only if

$$\frac{1}{2}(1+\sigma) < 1 - 2\varepsilon, \tag{12}$$

or

$$\varepsilon < \varepsilon_1 = \frac{1}{4}(1-\sigma). \tag{13}$$

As ε decreases past ε_1, the non-trivial anti-symmetric steady states bifurcate from $(0, 0)$.

3.1.2. *Two invariant sets.* Note that if $Q_1 = Q_2$ then $Q'_1 = Q'_2$, so that the set $Q_1 = Q_2$ in the (Q_1, Q_2) plane is invariant under the dynamical system. On this invariant set we have

$$Q' = (1-\kappa)Q + \kappa f(Q). \tag{14}$$

It is easy to show that $(0, 0)$ is unstable and $(-1, -1)$ and $(1, 1)$ are stable steady states of the dynamical system restricted to this set. Now note that if $Q_1 + Q_2 = 0$ then $Q'_1 + Q'_2 = 0$, so that the set $Q_1 + Q_2 = 0$ is also invariant under the dynamical system. On this set we have

$$Q' = (1-\kappa)Q + \kappa(1-2\varepsilon)f(Q). \tag{15}$$

The trivial steady state $Q = 0$ is stable as a solution of the dynamical system restricted to this set if $\varepsilon > \varepsilon_1$ but unstable if $\varepsilon < \varepsilon_1$, which is the condition for anti-symmetric steady states to exist. Each of the non-trivial anti-symmetric steady states is stable as solutions of this restricted dynamical system whenever the pair exists in $\mathbb{R}$.

3.1.3. *Stability of the symmetric and anti-symmetric steady states.* The results above give complete information on the stability of $(0, 0)$, showing that it has a single unstable eigenvalue (with eigenvector $(1, 1)^T$) if $\varepsilon > \varepsilon_1$, and two unstable eigenvalues (with eigenvectors $(1, 1)^T$ and $(1, -1)^T$) if $\varepsilon < \varepsilon_1$. To complete the analysis of the stability of the other symmetric and anti-symmetric steady states, we make use of the Jacobian matrix of the system in the (Q_1, Q_2) plane, given by

$$J(Q_1, Q_2) = \begin{pmatrix} 1 - \kappa + \kappa(1 - \varepsilon)f'(Q_1) & \varepsilon\kappa f'(Q_2) \\ \varepsilon\kappa f'(Q_1) & 1 - \kappa + \kappa(1 - \varepsilon)f'(Q_2) \end{pmatrix}. \tag{16}$$

Note that f is an odd function, so that f' is an even function, and so for the symmetric and anti-symmetric steady states, $f'(Q_1^*) = f'(Q_2^*)$. It follows that the eigenvectors at all of these steady states are $(1, 1)^T$ and $(1, -1)^T$, with eigenvalues

$$\lambda_+ = 1 - \kappa + \kappa f'(Q^*) \quad \text{and} \quad \lambda_- = 1 - \kappa + (1 - 2\varepsilon)\kappa f'(Q^*). \tag{17}$$

Clearly $0 < \lambda_- < \lambda_+$. For $Q^* = -1$ and $Q^* = 1$, $f'(Q^*) < 1$, so that $(-1, -1)$ and $(1, 1)$ have two stable eigenvalues, and hence are stable.

It remains to determine the stability of the anti-symmetric steady states. We know from subsection 3.1.2 that the condition that they are stable in the $(-1, 1)^T$ direction is identical to the condition that they exist in $\mathbb{R}$, $\varepsilon < \varepsilon_1$. The condition that they are stable in the $(1, 1)^T$ direction is that $f'(Q^*) < 1$. Since $f'(0) > 1$ and $f'(\pm 1) < 1$, and it is easy to show that $f''(Q)$ is positive for $Q \in (-1, 0)$ and negative for $Q \in (0, 1)$, this is true if Q^* is sufficiently close to ± 1. However $Q^* = \pm 1$ if $\varepsilon = 0$, and depends continuously on ε, so that these anti-symmetric steady states are stable for sufficiently small ε.

In fact, $(Q^*, -Q^*)$ (with $Q^* > 0$) is stable if $Q^* > \hat{Q}$, where $\hat{Q}$ is the positive solution of $f'(Q) = 1$. It is straightforward to show that

$$\hat{Q}^2 = \frac{-(2 + \sigma) + \sqrt{5 + 4\sigma}}{1 - \sigma}, \tag{18}$$

while we have already shown that

$$Q^{*2} = \frac{2(1 - 2\varepsilon) - (1 + \sigma)}{1 - \sigma}. \tag{19}$$

Hence $(Q^*, -Q^*)$ is stable whenever

$$2(1 - 2\varepsilon) > (1 + \sigma) - (2 + \sigma) + \sqrt{5 + 4\sigma}, \tag{20}$$

or

$$\varepsilon < \varepsilon_2 = \frac{1}{4}(3 - \sqrt{5 + 4\sigma}). \tag{21}$$

A calculation confirms that $\varepsilon_1 > \varepsilon_2 > 0$. As ε decreases towards zero, there are two bifurcation points, the one already discussed at $\varepsilon = \varepsilon_1 = \frac{1}{4}(1 - \sigma)$, where the pair of non-trivial anti-symmetric steady states bifurcate from $(0, 0)$, and one at $\varepsilon = \varepsilon_2 = \frac{1}{4}(3 - \sqrt{5 + 4\sigma})$, where a pair of steady states bifurcate from each of the anti-symmetric steady states, symmetrically placed with respect to the line

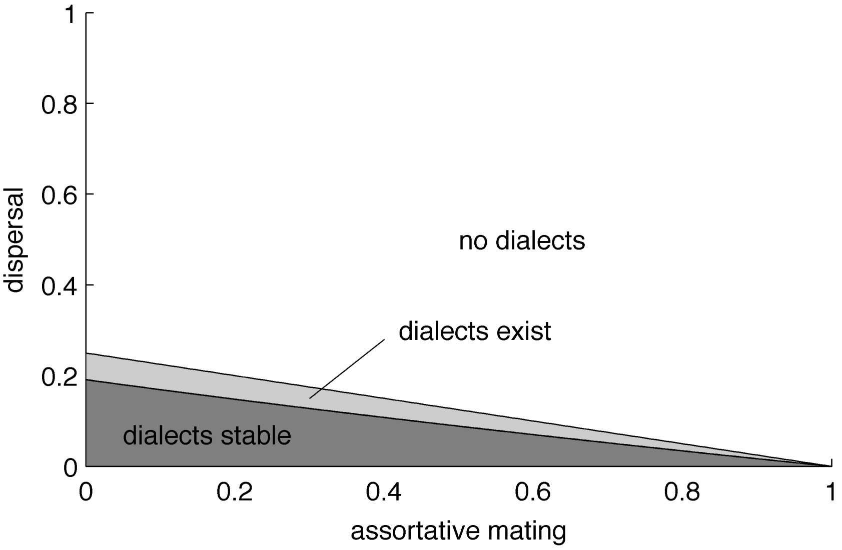

FIGURE 1. Bifurcation diagram for the post-dispersal learning model, indicating for which parameter values dialects exist or are indeed stable, summarised in inequalities (13) and (21).

$Q_1 + Q_2 = 0$. These new steady states are unstable, and their bifurcation confers stability on the anti-symmetric steady states. See Figure 1 for an illustration for which parameter values dialects are maintained.

3.2. **Conclusion.** We need both dispersal strength ε and assortative mating tendency σ to be sufficiently small in order for the anti-symmetric coexistence steady state to exist. Quantitatively, the requirement on ε and σ is given by inequality (13), $\varepsilon < \frac{1}{4}(1-\sigma)$. Dispersal must not be too great but assortative mating must be sufficiently strong in order to allow this. If ε is sufficiently small, the coexistence steady state will not only exist but will also be stable, as detailed in inequality (21). In this case, the system will tend to move towards a state in which both dialects exist; if the initial conditions are sufficiently close to this steady state, each dialect dominates in one of the sites. Note that the single-dialect steady states are also locally stable, and novel dialects cannot invade a one-dialect steady state.

Neither the adult mortality μ nor the learning parameter λ (nor therefore the product $\kappa = \mu(1-\lambda)$) are important in determining the existence or stability of the two-dialect steady state, although they will influence the speed at which any steady state is approached or left.

4. MODEL 2: PRE-DISPERSAL LEARNING

Model 1 assumes that juvenile birds first disperse and then learn their songs from conspecifics in their patch. Model 2 makes the opposite assumption.

Note that we do not model learning from parents here (which is already incorporated in the assortative mating), but from the local community, precisely as in

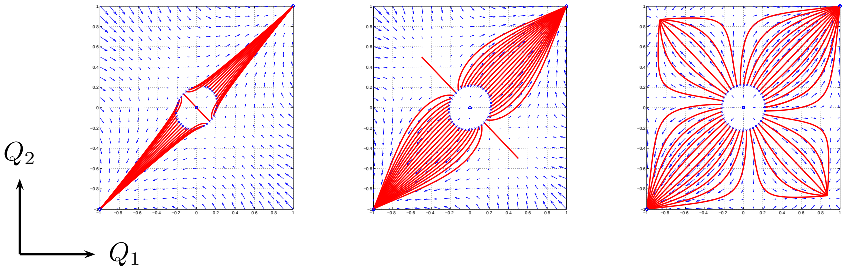

FIGURE 2. Example phase planes for system (8). Parameter values are $\lambda = 0.4, \mu = 0.4, \sigma = 0.2$ throughout, and $\varepsilon = 0.25$ (left), $\varepsilon = 0.15$ (middle) and $\varepsilon = 0.05$ (right). The critical ε-values at which bifurcations occur are $\varepsilon_1 = 0.2$ and $\varepsilon \approx 0.14$. As ε decreases, the nontrivial dialect steady states bifurcate from the origin, are first unstable (middle plot), and then become stable (right plot). Arrows indicate direction and speed of change of the dynamical system. The red lines are example solutions with initial conditions chosen on a circle centered around the origin. These plots illustrate how far the nontrivial steady states extend away from the origin before they stabilise, and also give some indication of the basins of attraction of these dialect steady states.

post-dispersal learning. The equations in Q_1 and Q_2 become subtly different:

$$Q_i' = (1-\mu)Q_i + \lambda\mu[(1-\varepsilon)Q_i + \varepsilon Q_j)] + \mu(1-\lambda)[(1-\varepsilon)q_i + \varepsilon q_j], \quad i,j = 1,2 \; i \neq j. \tag{22}$$

where, as before,

$$q_i = f(Q_i), \quad i = 1,2.$$

In the previous model, the εQ_2 term was εQ_1, so that we could write the sum $\lambda\mu((1-\varepsilon)Q_1 + \varepsilon Q_2))$ as $\lambda\mu Q_1$, and take $\kappa = \mu(1-\lambda)$ as a combined parameter. This is no longer the case.

The dynamics of this model are still invariant on the $Q_1 = Q_2$ line and on $Q_1 = -Q_2$. We again have the symmetric steady states $(-1,-1)$, $(0,0)$, and $(1,1,)$.

4.1. **Analysis.** The eigenvalues at the symmetric steady states and anti-symmetric steady states (where $f'(Q^*) = f'(-Q^*)$) are

$$\Lambda_+ = 1 - \mu(1-\lambda)(1 - f'(Q^*)),$$

and

$$\Lambda_- = 1 - \mu(1-\lambda)(1 - f'(Q^*)) + 2\mu\varepsilon(\lambda(f'(Q^*) - 1) - f'(Q^*)).$$

At the steady state $(1,1)$ and $(-1,-1)$, $f'(\pm 1) = \sigma$, and $\Lambda_- < \Lambda_+ < 1$, so that these symmetric steady states are always stable.

Assuming $Q_1 = Q_2$, the system reduces again to (14), and so the origin is unstable as before when we restrict ourselves to this line. There are again anti-symmetric steady states $(Q^*, -Q^*)$ if Q^* solves

$$Q^* = \frac{(1-\lambda)(1-2\varepsilon)}{(1-\lambda(1-2\varepsilon))} f(Q^*).$$

Conditions on λ, ε and σ to ensure the existence of the anti-symmetric steady state are now

$$\frac{1}{2}(1+\sigma) < \frac{(1-\lambda)(1-2\varepsilon)}{(1-\lambda(1-2\varepsilon))}.$$

Setting $S = \frac{1}{2} + \frac{1}{2}\sigma$ and $E = 1 - 2\varepsilon$, we require

$$\frac{(1-\lambda)E}{1-E\lambda} > S. \tag{23}$$

Note that $S > \frac{1}{2}$ for $\sigma \in (0,1)$, and $1 - \lambda > 0$ for $\lambda \in (0,1)$. The only possibility for (23) to occur is if both $E > 0$ and $1 - E\lambda > 0$ (the combination $E < 0$ and $1 - \lambda E < 0$ are excluded within these parameter ranges). This allows us to rewrite (23) as

$$\lambda < \frac{E-S}{E(1-S)},$$

where we have used that $1 - S > 0$ for all $\sigma \in (0,1)$. So in all, we require that

$$E > 0,\ E > S,\ 1-\lambda E > 0,\ \lambda < \frac{E-S}{E(1-S)}. \tag{24}$$

Written in the original parameters, the anti-symmetric nontrivial steady states exist when

$$\varepsilon < \frac{1}{2},\ \varepsilon < \frac{1}{4}(1-\sigma),\ \lambda < \frac{1-4\varepsilon-\sigma}{(1-2\varepsilon)(1-\sigma)},\ \lambda < \frac{1}{1-2\varepsilon}.$$

Note that $\frac{1}{4}(1-\sigma) < \frac{1}{2}$ and that

$$\frac{1-4\varepsilon-\sigma}{(1-2\varepsilon)(1-\sigma)} = \frac{1}{1-2\varepsilon} - \frac{4\varepsilon}{(1-2\varepsilon)(1-\sigma)} < \frac{1}{1-2\varepsilon},$$

so the above set of requirements reduces to

$$\varepsilon < \varepsilon_1 = \frac{1}{4}(1-\sigma),\ \lambda < \lambda_1 = \frac{1-4\varepsilon-\sigma}{(1-2\varepsilon)(1-\sigma)}. \tag{25}$$

These conditions are also the conditions for $(Q^*, -Q^*)$ to be stable in the $(1,-1)$ direction, so that existence of these equilibria coincides with the origin losing stability in this direction and the new anti-symmetric steady states taking over this stability. The eigenvalue in this eigendirection is greater than 1 iff

$$\lambda\sigma - 2\lambda\varepsilon\sigma - \sigma - \lambda + 2\varepsilon\lambda - 4\varepsilon + 1 > 0.$$

One can readily check that this is the same as (23).

As before, however, the condition that $(Q^*, -Q^*)$ is unstable in the $(1,1)$ direction is that $f'(Q^*) > 1$. Again, $f'(0) = 2/(1+\sigma) > 1$, so if Q^* is small enough, it is still unstable. We again look for the value $\hat{Q}$ such that $f'(Q) = 1$, which solves (18), as before, whilst Q^* solves

$${Q^*}^2 = \frac{2K-(1+\sigma)}{1-\sigma}, \quad \text{where} \quad K = \frac{2(1-\lambda)(1-2\varepsilon)}{(1-\lambda(1-2\varepsilon))}.$$

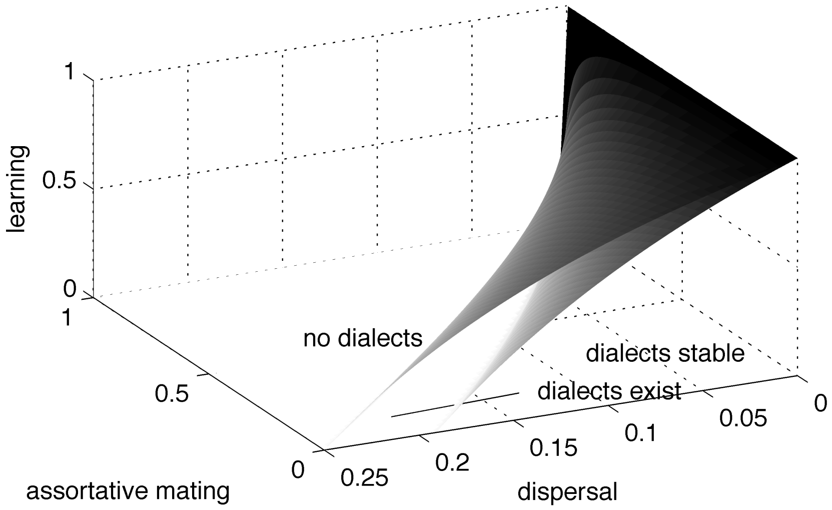

FIGURE 3. Bifurcation diagram for the pre-dispersal learning model, indication when dialect steady states exist or are indeed stable, summarised in inequalities (25) and (27).

Hence, $(Q^*, -Q^*)$ is stable whenever $Q^* > \hat{Q}$, i.e., whenever

$$K > S' := -\frac{1}{2} + \frac{1}{2}\sqrt{5+4\sigma}.$$

So we again have to solve

$$\frac{(1-\lambda)E}{1-E\lambda} > S'.$$

Since $S' > \frac{1}{2}$, and, as before, $S' \in (\frac{1}{2}, 1)$ for all $\sigma \in (0,1)$, we get analogous to (24),

$$E > 0,\ E > S',\ 1 - \lambda E > 0,\ \lambda < \frac{E - S'}{E(1-S')}. \tag{26}$$

It is straightforward to show that

$$\frac{E-S'}{E(1-S')} < \frac{1}{E},$$

so that the conditions, in the original parameters, reduce to

$$\varepsilon < \varepsilon_2 = \frac{1}{4}(3 - \sqrt{5+4\sigma}), \quad \lambda < \lambda_2 = \frac{3 - 4\varepsilon - \sqrt{5+4\sigma}}{(1-2\varepsilon)(3-\sqrt{5+4\sigma})}. \tag{27}$$

See Figure 3 for an illustration for which parameter values dialects are maintained, and some example phase planes in Figure 4.

4.2. **Conclusion.** Our second model leads to very similar results compared to the first, and are summed up in inequality (27): if dispersal ε and learning probability λ satisfy

$$\varepsilon < \varepsilon_2 = \frac{1}{4}(3 - \sqrt{5+4\sigma}), \quad \lambda < \lambda_2 = \frac{3 - 4\varepsilon - \sqrt{5+4\sigma}}{(1-2\varepsilon)(3-\sqrt{5+4\sigma})}$$

then dialects persist. The only real difference is that the learning parameter λ now does play a role in determining the occurrence of dialects arising from a well-mixed population—but a seemingly negative one. As with dispersal, learning must not be too strong for dialects to arise, which is opposite to our expectations. After all, imitative learning is always suggested to be a prerequisite for the emergence of dialectal patterns. In our second model, however, strong pre-dispersal learning will impede dialect formation and undermine dialect maintenance. This is contrary to biological understanding (see the Discussion), and suggests that a learning rule as implemented in this model is not realistic. Note also, that, as before, rare song types cannot invade a one-dialect steady state.

5. Model 3: Adaptive song types

Since post-dispersal learning is more commonly found than pre-dispersal learning, we go back to Model 1. Let us now further assume that in each patch the mortality rate depends on the song type, so that (4) becomes

$$P'_{ij} = (1-\mu_{ij})P_{ij} + (\mu_{ij}P_{ij} + \mu_{ik}P_{ik})p''_{ij} \tag{28}$$

for each $i = 1, 2$, $j = 1, 2$, where $k \neq j$. Note that all the adults who die in patch i each year are replaced by the same number of juveniles, as in the basic model, so that P_{ij} again represents a frequency, with $P_{i1} + P_{i2} = 1$ for each i. Let song type i be adapted to patch i, in such a way that, for simplicity, $\mu_{11} = \mu_{22} = \mu(1-\delta)$, $\mu_{12} = \mu_{21} = \mu(1+\delta)$, for some $\delta \in [0, 1)$. We define $Q_i = P_{i1} - P_{i2}$, as before, and we obtain

$$Q'_1 = Q_1 + \mu(\delta - Q_1) + \mu(1-\delta Q_1)\left\{\lambda Q_1 + (1-\lambda)(1-\varepsilon)f(Q_1) + (1-\lambda)\varepsilon f(Q_2)\right\},$$

$$Q'_2 = Q_2 - \mu(\delta + Q_2) + \mu(1+\delta Q_2)\left\{\lambda Q_2 + (1-\lambda)(1-\varepsilon)f(Q_2) + (1-\lambda)\varepsilon f(Q_1)\right\}.$$

Note that $(1, 1)$ and $(-1, -1)$ are still steady states of this system, but $(0, 0)$ is not.

Note also that if each song has a higher frequency of occurrence than the other where its mortality is lower (song 1 in patch 1 and song 2 in patch 2), then $Q_1 > 0$, $Q_2 < 0$, so we shall refer to the two-song steady state (Q_1^*, Q_2^*) as adaptive if $Q_1^* > 0$, $Q_2^* < 0$ and maladaptive if $Q_1^* < 0$, $Q_2^* > 0$.

It is easy to show that if $Q_1 + Q_2 = 0$ then $Q'_1 + Q'_2 = 0$, so that $Q_1 + Q_2 = 0$ is invariant, and $(Q_1^*, Q_2^*) = (Q^*, -Q^*)$ is an anti-symmetric steady state of the system as long as

$$Q^* = Q^* + \mu(\delta - Q^*) + \mu(1-\delta Q^*)\left\{\lambda Q^* + (1-\lambda)(1-2\varepsilon)f(Q^*)\right\}. \tag{29}$$

Writing $\tau = 2/(1-\sigma) = 1/(2\varepsilon_1)$, so that $f(Q) = \tau Q/(\tau - 1 + Q^2)$, this becomes

$$\lambda\delta Q^{*4} + (1-\lambda)Q^{*3} + a_2\delta Q^{*2} - (1-\lambda)(1-2\varepsilon\tau)Q^* - \delta(\tau-1) = 0, \tag{30}$$

where $a_2 = -(1+\lambda) + \tau(\lambda + (1-\lambda)(1-2\varepsilon))$. If $\delta = 0$ this quartic equation reduces to a cubic with solutions $Q^* = 0$ and $Q^* = \pm\sqrt{1-2\varepsilon\tau} = \pm\sqrt{1-\varepsilon/\varepsilon_1}$, the last two real if $\varepsilon < \varepsilon_1$. The model with $\delta = 0$ is the non-adaptive Model 1, and these are the steady states for Model 1, so this is as expected. These steady states are the ε-axis and the parabola sketched in figure 5.

For $\delta > 0$, it may be shown that the only relevant roots (i.e. real roots in $[-1, 1]$) of the quartic equation (30) are as sketched on the bifurcation diagram as a perturbation of the configuration for $\delta = 0$. For all values of ε there is a positive root $Q^* > 0$, (so that $Q_1^* > 0$, $Q_2^* < 0$, and the steady state is adaptive), which is always stable as a solution of the dynamical system restricted to $Q_1 + Q_2 = 0$, but stabilises

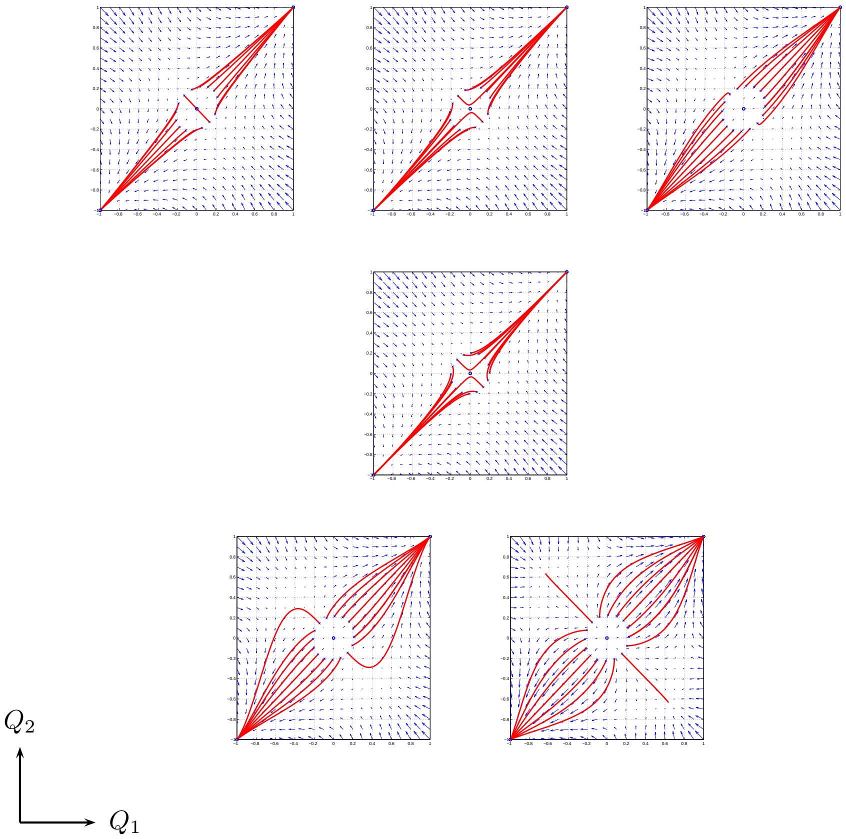

FIGURE 4. Example phase planes for the predisperal learning model (22). Parameter values are $\mu = 0.4, \sigma = 0.2$ throughout. In the top row, $\epsilon = 0.25$, so that $\lambda_2 < \lambda_1 < 0$ and anti-symmetric steady states never appear for any $\lambda > 0$, but they are unstable for all $\lambda > 0$. In these simulations, $\lambda = 0$ and $\lambda = 0.2$ for illustration. In the middle row, $\varepsilon = 0.18$, so that $\lambda_1 > 0$ but $\lambda_2 < 0$. In other words, anti-symmetric steady states do form for $\lambda < \lambda_1$ (in this illustration, $\lambda = 0.1 < 0.156 \equiv \lambda_1$), but they are never stable for any $\lambda > 0$. In the bottom row, $\varepsilon = 0.1$. For these values of ε and σ, $\lambda_1 \equiv= 0.625$, and $\lambda_2 = 0.405$. Hence, with sufficiently little learning, dialects are formed. From left to right in this row, $\lambda = 0.8, 0.5, 0.2$.

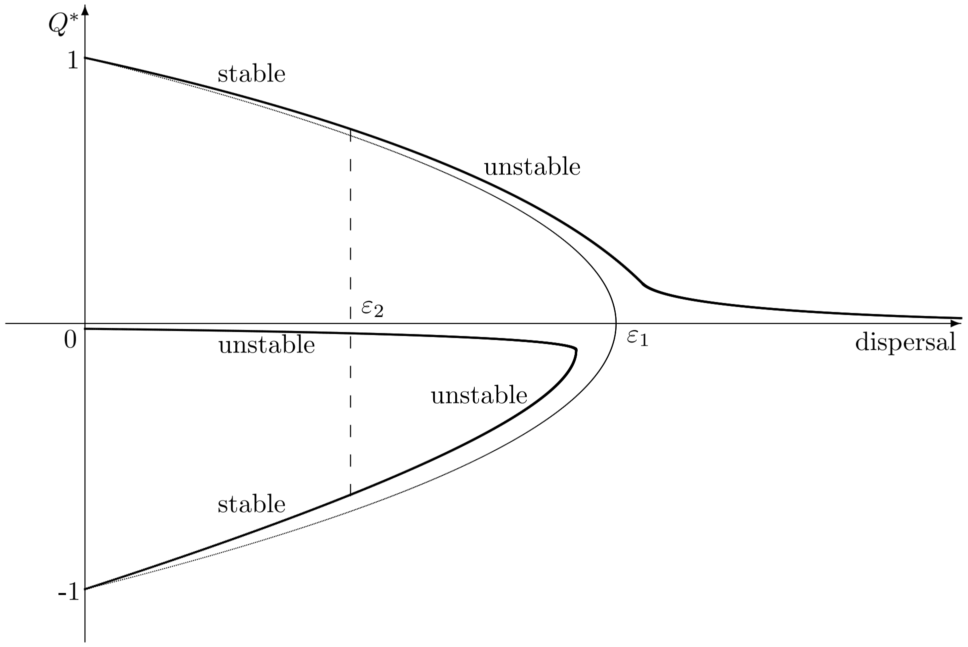

FIGURE 5. The bifurcation diagram for Model 3.

even without this restriction through a bifurcation near $\varepsilon = \varepsilon_2 = \frac{1}{4}(3 - \sqrt{5 + 4\sigma})$. In addition to this positive root of (30), there is a saddle-node bifurcation point at a value of $\varepsilon < \varepsilon_1 = \frac{1}{4}(1 - \sigma) = 1/(2\tau)$ where a pair of negative roots, corresponding to maladaptive steady states, appear. The root closer to $Q^* = 0$ is always unstable, while the more negative one is always stable as a solution of the dynamical system restricted to $Q_1 + Q_2 = 0$, but becomes unconditionally stable through a bifurcation near $\varepsilon = \varepsilon_2$ in a similar way to the positive root.

We have indicated that the adaptive two-song steady state is stable for ε sufficiently small, but we wish to know whether it attracts solutions initially close to one of the symmetric (single-song) steady states $(1, 1)$ and $(-1, -1)$, so that we might expect two songs to arise from one. In the case $\delta = 0$ the domains of attraction of the two anti-symmetric steady states expand at the expense of those of the symmetric steady states $(1, 1)$ and $(-1, -1)$ as ε decreases further, but do not reach these single-song steady states until ε reaches zero, (when the patches are completely isolated), so that for $\varepsilon > 0$ the single-song steady states are always stable. We now wish to investigate the stability of the single-song steady states $(1, 1)$ and $(-1, -1)$ for $\delta > 0$.

A straightforward calculation shows that the Jacobian matrix at $(1, 1)$ is given by

$$J(1,1) = \begin{pmatrix} 1 - A - \delta C & B - \delta D \\ B + \delta D & 1 - A + \delta C \end{pmatrix}, \tag{31}$$

where $A = \mu(1 - \lambda)(1 - (1 - \varepsilon)f'(1))$, $B = \varepsilon\mu(1 - \lambda)f'(1)$, $C = (1 + \lambda) + (1 - \lambda)(1 - \varepsilon)f'(1)$, and $D = \varepsilon(1 - \lambda)f'(1)$. The eigenvalues at $(1, 1)$ are therefore given by $\Lambda_\pm = 1 - A \pm \sqrt{B^2 + \delta^2(C^2 - D^2)}$. These are both stable for $\delta = 0$, but an eigenvalue passes through 1 as δ passes through δ_1, where $\delta_1^2 = \frac{A^2-B^2}{C^2-D^2}$. Noting that $f'(1) = \sigma$, we have

$$\delta_1^2 = (1-\lambda)^2 \frac{(1-(1-\varepsilon)\sigma)^2 - \varepsilon^2\sigma^2}{((1+\lambda)+(1-\lambda)(1-\varepsilon)\sigma)^2 - (1-\lambda)^2\varepsilon^2\sigma^2}. \tag{32}$$

or

$$\delta_1^2 = (1-\lambda)^2 \frac{(1-\sigma)(1-(1-2\varepsilon)\sigma)}{(1+\lambda)+(1-\lambda)\sigma)((1+\lambda)+(1-2\varepsilon)(1-\lambda)\sigma)} < 1. \tag{33}$$

A sufficient advantage $\delta > \delta_1$ to the favoured type destabilises $(1,1)$. Solutions of the dynamical system with initial conditions $(1,1)$ (or $(-1,-1)$) presumably then tend to the anti-symmetric steady state in the fourth quadrant as $t \to \infty$. See Figure 6 for some phase plane examples with increasing δ.

As δ increases, then (at least for ε small) the other anti-symmetric steady state remote from the origin remains stable until it hits the anti-symmetric steady state moving from the vicinity of the origin and disappears through a saddle-node bifurcation at $\delta = \delta_2$. At least for ε small, $\delta_2 < \delta_1$, so that this saddle-node bifurcation occurs before the bifurcation that destabilises $(1,1)$, and neither anti-symmetric steady state with $Q_1^* < 0$ is ever stable while $(1,1)$ is unstable.

5.1. **Conclusion.** Our third model, including environmental selection, yields dramatically different results compared to the first two models. Giving birds singing particular song types in a patch fitness benefits destroys the fully mixed steady state. Instead, we find an asymmetrical situation with a majority of birds singing the song type in the patch with highest fitness benefits. Sufficiently high penalties on singing the wrong song type in a patch (corresponding to $\delta > \delta_1$, where δ_1 is given by (33)) now does allow for new song types to invade. This does not occur in any of the other models.

6. MODEL 4: NONLINEAR FREQUENCY-DEPENDENT LEARNING

We go back to Model 1 since post-dispersal learning seems more common than pre-dispersal learning, as evidenced from well-studied species such as Bewicks wrens, *Thryomanes bewickii* (Kroodsma, 1974), white-crowned sparrows, *Zonotrichia leucophrys* (Baptista & Petrinovich, 1984; Nelson et al., 1996), great tits, *Parus major* (McGregor & Krebs, 1989), indigo buntings, *Passerina cyanea* (Payne, 1981), European starlings, *Sturnus vulgaris* (Mountjoy & Lemon, 1995), and canaries, *Serinus canarius* (Nottebohm & Nottebohm, 1978).

We thus consider the original post-dispersal Model 1, excluding assortative mating but assuming panmictic mating. It is still assumed that a young bird is more likely to learn the dialect that is sung by the majority of local adults, but the probability that it is learnt is now more than its majority share. This happens, for instance, if the young bird samples a finite number of local songs and chooses the most common song type in the sample.

The juvenile birds may now keep their natal dialects, with probability $1-\lambda$, or acquire a new dialect by a process of sampling from neighbouring adult birds, with probability λ. The sampling process is assumed to favour the dialect that is locally more common, in a stronger way than acquiring it from a single neighbour chosen at random. Let p''_{ij} be the frequency of juveniles at site i with dialect j after

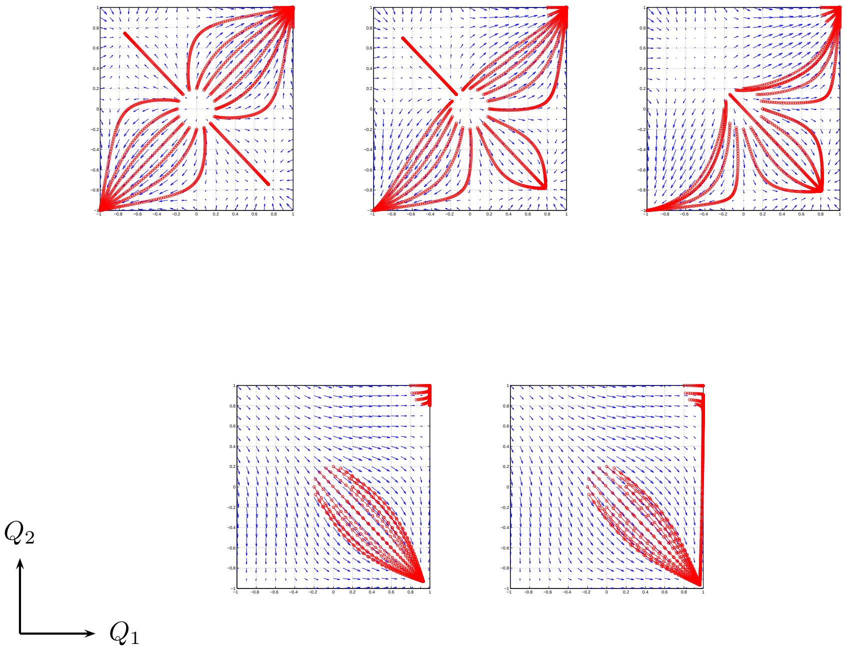

FIGURE 6. Example phase planes for system (5)–(5). Parameter values are $\lambda = 0.1, \mu = 0.1, \sigma = 0.1$, $\varepsilon = 0.1$ throughout, and $\delta = 0, 0.05, 0.1, 0.5, 0.7$ from top to bottom, left to right. The critical δ-value at which $(1,1)$ loses stability if $\delta_1 = 0.69$. Note that the solution tend to the anti-symmetric steady state in the fourth quadrant. The other anti-symmetrical steady state seems to remain stable, but ceases to exist when δ keeps increasing.

learning. Then

$$p''_{ij} = (1-\lambda)p'_{ij} + \lambda \frac{P^z_{ij}}{P^z_{ij} + P^z_{ik}}, \tag{34}$$

where $P_{ik} = 1 - P_{ij}$ is the frequency of adult birds in patch i singing the alternative dialect k. The parameter z measures how strongly the population favours the majority dialect; a value $z = 1$ gives $p''_{ij} = (1-\lambda)p'_{ij} + \lambda P_{ij}$, and is equivalent to sampling a single neighbour chosen at random, and we shall normally assume that $z > 1$.

Recall that a fraction μ of adults die each year and are replaced in the breeding population by the same number of one-year-old birds. Let P'_{ij} be the frequency of birds in patch i with dialect j at the subsequent census. Then

$$P'_{ij} = (1-\mu)P_{ij} + \mu p''_{ij}. \tag{35}$$

for each $i = 1, 2$, $j = 1, 2$.

6.1. **Analysis.** Note that P_{ij}, p_{ij}, p'_{ij} and p''_{ij} are frequencies (with respect to j), so that we need only consider new variables $Q_i = P_{i1} - P_{i2}$, $q_i = p_{i1} - p_{i2}$, etc. Then, from (35),

$$Q'_i = (1-\mu\lambda)Q_i - \mu(1-\lambda)\varepsilon Q_i + \mu(1-\lambda)\varepsilon Q_j + \mu\lambda g(Q_i), \tag{36}$$

where $j \neq i$ and

$$g(Q) = \frac{(1+Q)^z - (1-Q)^z}{(1+Q)^z + (1-Q)^z}. \tag{37}$$

Note that g is an odd function, and satisfies

$$g(0) = 0, \quad g(1) = 1, \quad g(-1) = -1. \tag{38}$$

Its derivative is given by

$$g'(Q) = \frac{4z(1-Q^2)^{z-1}}{((1+Q)^z + (1-Q)^z)^2}, \tag{39}$$

which is of course an even function, and (if $z > 1$) satisfies

$$g'(0) = z, \quad g'(1) = 0, \quad g'(-1) = 0. \tag{40}$$

It is also easy to show that (if $z > 1$) the second derivative g'' satisfies

$$g''(Q) < 0 \ \text{ for } 0 < Q < 1, \quad g''(Q) > 0 \ \text{ for } -1 < Q < 0. \tag{41}$$

6.1.1. *Symmetric and anti-symmetric steady states.* Since $g(0) = 0$, $g(1) = 1$, $g(-1) = -1$, there are three steady states that are symmetric under the transformation $(Q_1, Q_2) \to (Q_2, Q_1)$, $Q_1 = Q_2 = 1$ and $Q_1 = Q_2 = -1$, corresponding to extinction of dialect 2 and 1 respectively, and $Q_1 = Q_2 = 0$, corresponding to equal numbers of each dialect in each site, $P_{11} = P_{12} = P_{21} = P_{22} = \frac{1}{2}$. The points $(Q^*, -Q^*)$ and $(-Q^*, Q^*)$ are non-trivial anti-symmetric steady states of the system (under the same transformation) if

$$(\lambda - 2(1-\lambda)\varepsilon)Q^* = \lambda g(Q^*). \tag{42}$$

Let us consider λ as a bifurcation parameter, and look for such anti-symmetric solutions bifurcating from the trivial solution. They do so where

$$(\lambda - 2(1-\lambda)\varepsilon) = \lambda g'(0) = \lambda z, \tag{43}$$

or

$$\lambda = \lambda_1 = \frac{2\varepsilon}{z - 1 + 2\varepsilon}. \tag{44}$$

As λ increases past λ_1, two non-trivial anti-symmetric steady states bifurcate from $(0,0)$.

For $z = 1$, $\lambda_1 = 1$, and there is no bifurcation for realistic values of λ; we have already stated that we expect $z > 1$ for pattern formation. For $z = 2$, we can find the anti-symmetric steady states explicitly, since then $g(Q) = 2Q/(1+Q^2)$. They are given by

$$Q^* = \pm\sqrt{\frac{\lambda - 2(1-\lambda)\varepsilon}{\lambda + 2(1-\lambda)\varepsilon}}. \tag{45}$$

6.1.2. *Two invariant sets.* Note that if $Q_1 = Q_2$ then $Q_1' = Q_2'$, so that the set $Q_1 = Q_2$ in the (Q_1, Q_2) plane is invariant under the dynamical system. On this invariant set we have

$$Q' = (1-\mu\lambda)Q + \mu\lambda g(Q). \tag{46}$$

Since $g'(0) = z > 1$, and $g'(-1) = g'(1) = 0$ for $z > 1$, then $(0,0)$ is unstable and $(-1,-1)$ and $(1,1)$ are stable steady states of the dynamical system restricted to this set. Now note that if $Q_1 + Q_2 = 0$ then $Q_1' + Q_2' = 0$, since g is odd, so that the set $Q_1 + Q_2 = 0$ is also invariant under the dynamical system. On this set we have

$$Q' = (1-\mu\lambda)Q - 2\mu(1-\lambda)\varepsilon Q + \mu\lambda g(Q). \tag{47}$$

The trivial steady state $Q = 0$ is stable as a solution of the dynamical system restricted to this set if $\lambda < \lambda_1$ but unstable if $\lambda > \lambda_1$, which is the condition for anti-symmetric steady states to exist. Each of the non-trivial anti-symmetric steady states is stable as solutions of this restricted dynamical system whenever the pair exists in $\mathbb{R}$.

6.1.3. *Stability of the symmetric and anti-symmetric steady states.* The results above give complete information on the stability of $(0,0)$, showing that it has a single unstable eigenvalue (with eigenvector $(1,1)^T$) if $\lambda < \lambda_1$, and two unstable eigenvalues (with eigenvectors $(1,1)^T$ and $(1,-1)^T$) if $\lambda > \lambda_1$. To complete the analysis of the stability of the other symmetric and anti-symmetric steady states, we make use of the Jacobian matrix of the system in the (Q_1, Q_2) plane, given by

$$J(Q_1,Q_2) = \begin{pmatrix} 1-\mu\lambda-\mu(1-\lambda)\varepsilon+\mu\lambda g'(Q_1) & \mu(1-\lambda)\varepsilon \\ \mu(1-\lambda)\varepsilon & 1-\mu\lambda-\mu(1-\lambda)\varepsilon+\mu\lambda g'(Q_2) \end{pmatrix}. \tag{48}$$

Since g' is an even function, then $g'(Q_1^*) = g'(Q_2^*)$ for the symmetric and anti-symmetric steady states. It follows that the eigenvectors at all of these steady states are $(1,-1)^T$ and $(1,1)^T$, with eigenvalues

$$\Lambda_- = 1-\mu\lambda-2\mu(1-\lambda)\varepsilon+\mu\lambda g'(Q^*) \quad \text{and} \quad \Lambda_+ = 1-\mu\lambda+\mu\lambda g'(Q^*). \tag{49}$$

Clearly $0 < \Lambda_- < \Lambda_+$, for realistic values of $\varepsilon < \frac{1}{2}$. For $Q^* = -1$ and $Q^* = 1$, $g'(Q^*) = 0 < 1$, so that $(-1,-1)$ and $(1,1)$ have two stable eigenvalues, and hence are stable.

It remains to determine the stability of the anti-symmetric steady states. We know from subsection 6.1.2 that the condition that they are stable in the $(1,-1)^T$ direction is identical to the condition that they exist in $\mathbb{R}$, $\lambda > \lambda_1$. The condition that they are stable in the $(1,1)^T$ direction is that $g'(Q^*) < 1$. Since $g'(0) = z > 1$ and $g'(\pm 1) = 0 < 1$, and $g''(Q)$ is negative for $Q \in (-1,0)$ and positive for $Q \in$

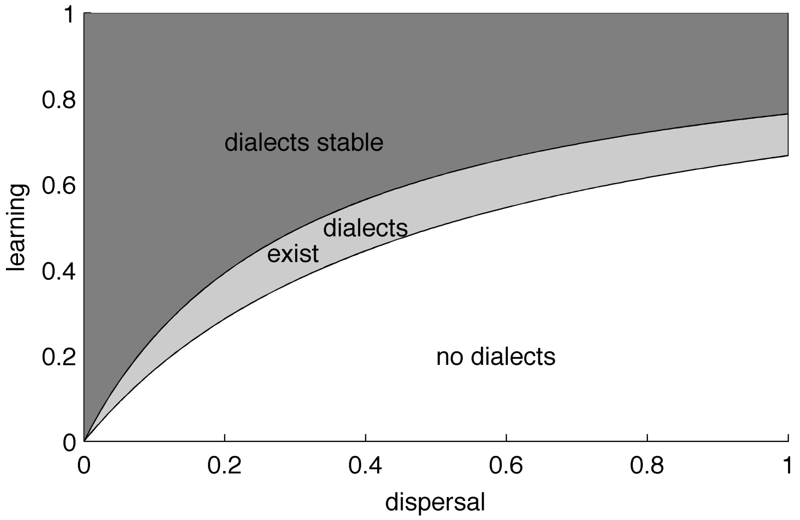

FIGURE 7. Bifurcation diagram of the nonlinear frequency-dependent learning model, indicating for which parameter values dialects exist or are indeed stable. The dark-gray area indicates the region of stable dialects for the choice $z = 2$, see inequality (53).

$(0, 1)$, this is true if Q^* is sufficiently close to ± 1. However $Q^* = \pm 1$ if $\lambda = 1$, and depends continuously on λ, so that these anti-symmetric steady states are stable for λ sufficiently close to unity.

In fact, $(Q^*, -Q^*)$ (with $Q^* > 0$) is stable if $Q^* > \hat{Q}$, where $\hat{Q}$ is the positive solution of $g'(Q) = 1$. For $z = 2$ it is straightforward to show that

$$\hat{Q}^2 = -2 + \sqrt{5}, \tag{50}$$

while we have already shown that

$$Q^{*2} = \frac{\lambda - 2(1-\lambda)\varepsilon}{\lambda + 2(1-\lambda)\varepsilon}. \tag{51}$$

Hence $(Q^*, -Q^*)$ is stable whenever

$$\lambda - 2(1-\lambda)\varepsilon > (\lambda + 2(1-\lambda)\varepsilon)(-2 + \sqrt{5}), \tag{52}$$

or

$$\lambda > \lambda_2 = \frac{2\varepsilon(\sqrt{5} - 1)}{(3 - \sqrt{5}) + 2\varepsilon(\sqrt{5} - 1)}. \tag{53}$$

A calculation confirms that $0 < \lambda_1 < \lambda_2 < 1$. As λ increases towards unity, there are two bifurcation points, the one already discussed at $\lambda = \lambda_1 = 2\varepsilon/(1+2\varepsilon)$ (for $z = 2$), where the pair of non-trivial anti-symmetric steady states bifurcates from $(0, 0)$, and one at $\lambda = \lambda_2 = 2\varepsilon(\sqrt{5} - 1)/(3 - \sqrt{5} + 2\varepsilon(\sqrt{5} - 1))$, where a pair of steady states bifurcates from each of the anti-symmetric steady states, symmetrically placed with respect to the line $Q_1 + Q_2 = 0$. These new steady states are unstable, and their bifurcation confers stability on the anti-symmetric steady states. See Figure 7 for a bifurcation diagram illustrating for which parameter values dialects are maintained.

6.2. **Conclusion.** The fourth and last model confirms, using a more realistic learning mode, the importance of dispersal and the learning factor in the persistence of two different dialects. For fixed dispersal strength ε, we need learning probability λ to be sufficiently large, as detailed in inequality (44), in order for the anti-symmetric coexistence steady states to exist. That is, dispersal must not be too great but learning must be sufficiently strong and sufficiently favourable to the majority dialect in order to allow this. If λ is sufficiently large, see inequality (53), the coexistence steady states will not only exist but will also be stable. Now the system will tend to a state in which both dialects exist, and each dialect dominates in one of the sites, if the initial conditions are sufficiently close to this steady state. Note that the single-dialect steady states are again locally stable, so that this does not explain how such a two-dialect steady state may be arrived at from a one-dialect ancestral state.

Although we have made a dramatic change in the learning component of this last model, the adult mortality μ remains unimportant in determining the existence or stability of the two-dialect steady state. It will of course influence the speed at which any steady state is approached or left.

7. Discussion

Our theoretical explorations confirm previous understanding but also reveal several new insights. First, as expected, the combination of low dispersal and strong assortative mating based on song promotes dialect maintenance (models 1 to 4). Second, although variation in post-dispersal learning propensity affects the rate of change, a simple linear rule of tutor frequency-dependent copying probability has little effect on the probability of maintenance for our two neighboring song dialects (model 1). Third, variation in pre-dispersal learning propensity does affect the maintenance of dialect boundaries, but plays an undermining role if songs are largely and accurately copied from natal territories (model 2). Fourth, we revealed a strong environmental impact on dialectal maintenance patterns: the probability that rare song types can establish and remain depends heavily on whether there are fitness consequences to using a song type in a particular patch (model 3). And fifth, an impact of post-dispersal learning flexibility on dialect maintenance turns out to be dependent on the rule for tutor selection (model 4): dialect maintenance is enhanced by post-dispersal learning if dispersing males follow a non-linear rule and learn the most common local dialect with a greater probability than the fraction of local birds singing it.

7.1. **The impact of pre- and post-dispersal learning.** Pre-dispersal learning will promote acoustic convergence and merging of dialects. In such cases dialect maintenance will rely solely on limited dispersal and strong assortative mating. In isolated or small, newly colonized populations may still diverge by founder effects or the accumulation of random copy errors (Baker et al., 1987; Slabbekoorn & Peet, 2003). It is difficult to empirically verify the predictions of our model as pre-dispersal learning may be rare and restricted to species such as zebra finches, *Taeniopygia guttata* (Zann, 1997) and Darwin's finches, *Geospiza* sp. (Grant & Grant, 1996). However, the impact of strong pre-dispersal learning may yield patterns similar to those of non-learning species, as suggested by the model of Ellers & Slabbekoorn (2003). The available empirical examples of geographic patterns in

Model	**Parameters changed**	**Dialect maintenance**	**Dialect invasion**
1: post-dispersal, linear learning	ε (dispersal)	$-$	X
	λ (learning probability)	X	X
	μ (mortality)	X	X
	σ (assortative mating)	$+$	X
2: pre-dispersal, linear learning	ε	$-$	X
	λ	X	X
	μ	X	X
	σ	$+$	X
3: post-dispersal, adaptive song types	δ (measure of difference in mortality rates)	$+$	$+$
4: post-dispersal, nonlinear learning	ε	$-$	X
	λ	$+$	X
	(μ)	(X)	(X)
	(σ)	$(+)$	(X)

TABLE 1. Overview of the four different models considered, and the different outcomes when varying parameters. A +/- indicates that increasing the parameter has a positive(negative) influence on dialect maintenance or dialect invasion (i.e., establishment of new dialects); an X means changing the parameter has no effect. In model 3, parameters ε, λ, μ and σ were kept fixed. Parameter δ did not feature in models 1, 2 and 4. In model 4, only the effect of ε and λ were explicitly studied. The influence of parameters μ and σ is in all likelihood the same as in model 1, as indicated by the parentheses. "Linear (or nonlinear) learning" refers to "linear (or nonlinear) frequency-dependent learning", see Model 1 for additional explanation.

non-learning species indeed suggest a lack of discrete boundaries such as those expected with a dominant merging role for gene or meme flow (Appleby & Redpath, 1997; Peake & McGregror, 1999).

Post-dispersal learning has the potential to promote acoustic divergence, but dialect maintenance depends on the behavioural rule of tutor selection. The difference between our first and fourth model was the mode of learning. In model 1 we assumed linear frequency-dependent oblique learning. This rule implies that a dispersing male bird exposed to four different tutors, three singing one song types and one singing another, chooses the most common onewith probability 75%. In model 4, with the same set of tutors, the likelihood of learning the majority dialect is greater than 75%. In the latter condition, learning propensity does not have a neutral role in dialect maintenance: dialects are actively reinforced instead, and are easier to maintain at the same levels of dispersal and assortative mating.

A simple kind of voting mechanism, in which a bird chooses ('votes for') the most common type of song heard over a limited period of time, would give rise to this. Linear frequency-dependent learning would then be voting based on just one random sample from the population, and a nonlinear dependence would be created by voting after hearing more than one randomly chosen song. This suggests that a nonlinear rule should be the norm, not the exception. Simple voting mechanisms are known to cause patterns in a wide variety of biological systems (see e.g., Camazine et al. 2001). An alternative mechanism leading to a nonlinear frequency response might be selective attrition, in which birds learn a series of different songs, then subsequently lose most of these (Marler & Peters, 1982; Nelson, 1992).

7.2. **Environmental impact on dialectal maintenance and emergence.** Assuming direct fitness consequences for using a particular song type in a focal patch has a great impact on dialect maintenance, and indeed on dialect emergence. Among the models considered here, rare song types are never invasive and hence never remain, unless there are fitness benefits for using that rare song type. Extant patterns of geographic song variation may thus be caused (or promoted) initially by environmental impact, but may subsequently be maintained by a combination of low dispersal, strong assortative mating and a nonlinear rule for tutor selection.

It is currently difficult to assess the importance of these model results using empirical evidence. Studies are still mostly divided into those with a focus either on dialect or on environmental impact and change in acoustic structure (Slabbekoorn & Smith, 2002). There are, however, some interesting exceptions. The study on white-crowned sparrows in the Sierra Nevada mountains by MacDougall-Shackleton et al. (2002) indicates that males singing the local dialect may benefit from local adaptation, as they had higher fertility rates than males singing a deviating song that may have come from elsewhere. In another study, Luther & Baptista (2010) show dialectal changes in time in urban white-crowned sparrows of San Francisco, which may be driven by environmental advantages to high-frequency song variants in neighbourhoods where low-frequency traffic noise has increased.

Similar noise-dependent and discrete song variation can be found beyond the intra-specific level among different subspecies or closely-related species that live in different natural habitats. A clear example involves high-elevation and low-elevation woodwren populations living on the steep slopes of the South-American Andes (Dingle et al., 2008, 2010). In this model system, empirical data involve adjacent populations, even including a narrow zone with adjacent territories, where song variation remains as a distinct acoustic boundary. At low elevations there is much more high-frequency noise than at high elevations, probably mainly due to variation in the local insect communities, and at high elevations woodwrens sing more high-frequency song notes than at low elevations. Detailed song analyses and genetic determination of parental species and hybrids at the contact zone indicates the existence of mixed singers and even some individuals singing song repertoires which are entirely like those of the other population (Halfwerk et al. unpublished data). So, this is a case of environmental selection on song variation in which a distinct acoustic boundary is maintained despite proven capabilities of heterotypic song-learning.

7.3. **Theoretical work.** Several related theoretical studies on bird-song dialects focus mainly on the effect of social learning or dialectal patterns on biological processes (e.g., Ellers & Slabbekoorn, 2003; Lachlan & Servedio, 2004; Lachlan et al., 2004). Most of these models are spatially explicit and all are simulation models. However, the precise mechanisms with which dispersal, learning, or territorial defence and mate choice are assumed to work differ substantially between them. One common choice, however, is that song-learning is taken to be linearly frequency-dependent, as in our models 1 to 3.

A few studies, however, focus on the pattern formation itself. In an early model of repertoire maintenance, Williams & Slater (1990) implemented a linear oblique learning mode for birds in a spatial grid. Depending on the number of tutors from which a young bird could learn its repertoire, song types within these repertoires persisted for longer when song was learnt from more tutors. In Ellers & Slabbekoorn (2003), there is a mixed choice of predefined subpopulations from which song tutors are chosen, which ignores those subpopulation boundaries. Dialects do form in this model, contrary to our results from models 1–3. This indicates that biased learning within local populations and dispersal through assortative territory-holding potential can have a comparable impact on geographic variation in song as assortative mating in our current models.

In our models we assumed the occurrence of nonlinear frequency-dependent learning from song tutors. Some evolutionary aspects of this phenomenon, also termed conformity bias (Boyd & Richardson, 1985) have been modelled by Lachlan et al. (2004) using a game-theoretical approach. Our model 4 also suggests that a combination of conformity bias and assortative mating would already suffice to cause between-group differences, but it may even be simpler. Strigul (2009) recently showed that simple implementations of nonlinear frequency-dependent learning may cause emergence of homogeneous subgroups. In this model, individuals could copy traits from conspecifics in such a way that an agent was more likely to copy the trait if there was an overall greater match over all shared traits. In the models studied by Strigul, only imitation was considered, but other ingredients such as dispersal, mortality and assortative mating were left out. Self-organization into dialect groups was found to depend strongly on the mode of learning. Our models suggest that this cohesive force persists even in the face of dispersal and mortality.

7.4. **Outlook.** Much of the current literature presumes that the formation and maintenance of dialects is self-evident. Kroodsma (2004), in a recent overview chapter, wrote that "Dialects happen because birds tend to breed and therefore to sing in the same local area where they learned their songs." Our models show that even in the minimal case of two patches and two song types, the integrated outcome of the various mechanisms at play is not evident. Song-learning, when implemented as linearly frequency-dependent (the most common model choice), does not promote dialects in our models.

Although our current paper focused on dialect maintenance, it also reflects to some extent on dialect formation. It is most often thought that new dialects arise through founder effects (Kroodsma, 2004). This will undoubtedly be true in some cases. However, we believe that the nonlinear effects of assortative mating and song-learning, as implemented for instance in Model 4, could also give rise to the emergence of dialects when explored in a spatial model. Spatial pattern formation is one of the main fields in applied mathematics these last decades (Murray, 2003), and patterns can often be explained using very simple nonlinear mechanisms. There is still much scope for future work on birdsong dialects, in which the vast literature on song-learning, dispersal, assortative mating and habitat is taken into account more fully in more realistic models, to give rise to a coherent theory of birdsong dialects.

References

Appleby, B. M. & Redpath, S. M. (1997). Indicators of male quality in the hoots of tawny owls (*Strix aluco*). *J. Raptor Research*, 31, 65–70.

Baker, M. C. (2003). Local similarity and geographic differences in a contact call of the galah (*Cacatua roseicapilla assimilis*) in Western Australia. *Emu*, 103, 233–237.

Baker, M. C. & Cunningham, M. A. (1985). The biology of bird-song dialects. *Behav. Brain Sci.*, 8, 85–133.

Baker, M. C., Spitler-Nabors, K. J., Thompson, A. D. J., & Cunningham, M. A. (1987). Reproductive behaviour of female white-crowned sparrows: Effect of dialects and synthetic hybrid songs. *Animal Behaviour*, 35, 1766–1774.

Baptista, L. F., Boarman, W. I., & Kandianidis, P. (1983). Behavior and taxonomic status of Grayson's dove. *Auk*, 100, 907–919.

Baptista, L. F. & King, J. R. (1980). Geographical variation in song and song dialects of the montane white-crowned sparrow. *Condor*, 82, 267–284.

Baptista, L. F. & Petrinovich, L. (1984). Social interaction, sensitive phases and the song template hypothesis in the white-crowned sparrow. *Anim. Behav.*, 32, 172–181.

Bell, D. A., Trail, P. W., & Baptista, L. F. (1998). Song learning and vocal tradition in Nuttall's white-crowned sparrows. *Anim. Behav.*, 55, 939–956.

Borror, D. J. (1956). Variation in Carolina Wren songs. *Auk*, 73, 211–229.

Boyd, R. & Richardson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: Chicago University Press.

Camazine, S., Deneubourg, J.-L., Franks, N. R., Sneyd, J., Theraulaz, G., & Bonabeau, E. (2001). *Self-organization in biological systems*. Princeton University Press.

Catchpole, C. & Slater, B. (1995). *Bird Song - Biological Themes and Variations*. Cambridge University Press.

Cunningham, M. A., Baker, M. C., & Boardman, T. J. (1989). Microgeographic song variation in the Nuttall's white-crowned sparrow. *Condor*, 89, 261–275.

Derryberry, E. P. (2011). Male response to historical and geographical variation in bird song. *Biol. Letters*, 7, 57–59.

Dingle, C., Halfwerk, W., & Slabbekoorn, H. (2008). Habitat-dependent song divergence at subspecies level in grey-breasted wood-wren. *J. Evol. Biol.*, 21, 1079–1089.

Dingle, C., Poelstra, J. W., Halfwerk, W., Brinkhuizen, D. M., & Slabbekoorn, H. (2010). Asymmetric response patterns to subspeciesspecific song differences in allopatry and parapatry in the gray-breasted wood-wren. *Evolution*, 64, 3537–3548.

Edwards, S. V., Kingan, S. B., Calkins, J. D., Balakrishnan, C. N., Jennings, W. B., Swanson, W. J., & Sorenson, M. D. (2005). Speciation in birds: genes, geography, and sexual selection. *Proc. Nat. Acad. Sciences*, 102, 6550–6557.

Ellers, J. & Slabbekoorn, H. (2003). Song divergence and male dispersal among bird populations: a spatially explicit model testing the role of vocal learning. *Anim. Behav.*, 65, 671–681.

Gaunt, S. L. L., Baptista, L. F., Sánchez, J. E., & Hernández, D. (1994). Song learning as evidenced from song sharing in two hummingbird species (*Colibri coruscans* and *C. thalassinus*). *Auk*, 111, 87–103.

Goldstein, R. B. (1978). Geographic variation in Hoy call of Bobwhite *Colinus virginianus*. *Auk*, 95, 85–94.

Grant, B. R. & Grant, P. R. (1996). Cultural inheritance of song and its role in the evolution of Darwinś finches. *Evolution*, 50, 2471–2487.

Handford, P. (1988). Trill rate dialects in the rufous-collared sparrow, *Zonotrichia capensis*, in north-western Argentina. *Can. J. Zool.*, 66, 26582670.

Harbison, H., Nelson, D. A., & Hahn, T. P. (1999). Longterm persistence of song dialects in the mountain whitecrowned sparrow. *Condor*, 101, 133–148.

Kroodmsa, D. E. (1985). Goegraphic variation in songs of the Bewicks Wren complex - a search for correlations with avifaunal complexity. *Behav. Ecol. Sociobiol.*, 16, 143–150.

Kroodsma, D. E. (1974). Song learning, dialects and dispersal in the Bewicks wren. *Z. für Tierpsychologie*, 35, 352–380.

Kroodsma, D. E. (2004). The diversity and plasticity of birdsong. In P. Marler & H. Slabbekoorn (Eds.), *Nature's Music* (pp. 108–131). Elsevier Academic Press.

Kroodsma, D. E., Liu, W. C., Goodwin, E., & Bedell, P. A. (1999). The ecology of song improvisation as illustrated by North American Sedge Wrens. *Auk*, 116, 373–386.

Lachlan, R. & Slater, P. J. B. (1999). The maintenance of vocal learning by gene-culture interaction: the cultural trap hypothesis. *Proc. Roy. Soc. London B*, 266, 701–706.

Lachlan, R. F., Janik, V. M., & Slater, P. J. B. (2004). The evolution of conformity-enforcing behaviour in cultural communication systems. *Anim. Behav.*, 68, 561–570.

Lachlan, R. F. & Servedio, M. R. (2004). Song learning accelerates allopatric speciation. *Evolution*, 58(9), 2049–2063.

Lachlan, R. F. & Slater, P. J. B. (2003). Song learning by chaffinches: how accurate, and from where? *Anim. Behav.*, 65, 957–969.

Luther, D. & Baptista, L. F. (2010). Urban noise and the cultural evolution of bird songs. *Proc. Roy. Soc. London B*, 277, 469–473.

MacDougall-Shackleton, E. A., Derryberry, E. P., & Hahn, T. P. (2002). Nonlocal male mountain white-crowned sparrows have lower paternity and higher parasite loads than males singing local dialect. *Behav. Ecol.*, 13, 682–689.

Marler, P. (1970). A comparative approach to vocal learning: song development in white-crowned sparrows. *J. Comp. Physiol. Psychol.*, 71, 1–25.

Marler, P. & Peters, S. (1982). Developmental overproduction and selective attrition: New processes in the epigenesis of birdsong. *Devel. Psychiobiol.*, 15, 369–378.

Marler, P. & Slabbekoorn, H. (2004). *Nature's Music.* Elsevier Academic Press.

Marler, P. & Tamura, M. (1962). Song "dialects" in three populations of white-crowned sparrows. *Condor*, 64, 368–377.

McGregor, P. K. & Krebs, J. R. (1989). Song learning in adult great tits (*Parus major*): effect of neighbours. *Behaviour*, 108, 139–159.

Mountjoy, D. E. & Lemon, R. E. (1995). Extended song learning in wild European starlings. *Anim. Behaviour*, 49, 357–366.

Mundinger, P. C. (1982). Microgeographic and macrogeographic variation in the acquired vocalizations of birds. In D. E. Kroodsma, E. H. Miller, & H. Quellet (Eds.), *Acoustic Communication in Birds, vol. I* (pp. 147–208). New York: Academic Press.

Murray, J. D. (2003). *Mathematical Biology II: Spatial Models and Biomedical Applications*. Springer-Verlag.

Nelson, D. A. (1992). Song overproduction and selective attrition lead to song sharing in the field sparrow (*Spizella pusilla*). *Behav. Ecol. Sociobiol.*, 30, 415–424.

Nelson, D. A. (2000). Song overproduction, selective attrition, and song dialects in the white-crowned sparrow. *Anim. Behav.*, 60, 887–898.

Nelson, D. A., Hallberg, K. I., & Soha, J. A. (2004). Cultural evolution of Puget Sound white-crowned sparrow song dialects. *Ethology*, 110, 879–908.

Nelson, D. A., Khanna, H., & Marler, P. (2001). Learning by instruction or selection: Implications for patterns of geographic variation in bird song. *Behaviour*, 138, 1137–1160.

Nelson, D. A., Marler, P., & Morton, M. L. (1996). Overproduction in song development: an evolutionary correlate with migration. *Anim. Behaviour*, 51, 1127–1140.

Nottebohm, F. & Nottebohm, M. E. (1978). Relationship between song repertoire and age in the canary, s*erinus canarius*. *Z. für Tierpsychologie*, 46, 298–305.

Payne, R. B. (1978). Microgeographic variation in songs of the splendid sunbird *Nectarinia coccinigaster*: population phenetics, habitats, and song dialects. *Behaviour*, 65, 282–308.

Payne, R. B. (1981). Song learning and social interaction in indigo buntings. *Anim. Behaviour*, 29, 688–697.

Peake, T. M. & McGregror, P. K. (1999). Geographical variation in the vocalization of the corn crake *Crex crex*. *Ethol. Ecol. Evol.*, 11, 123–137.

Podos, J. & Warren, P. S. (2007). The evolution of geographic variation in birdsong. *Adv. Study of Behav.*, 37, 403–458.

Price, T. (2008). *Speciation in Birds*. Roberts and Company, Greenwood Village.

Rivera-Gutierrez, H. F., Matthysen, E., Adriaensen, F., & Slabbekoorn, H. (2010). Repertoire sharing and song similarity between great tit males decline with distance between forest fragments. *Ethology*, 116, 951–960.

Slabbekoorn, H., Jesse, A., & Bell, D. (2003). Microgeographic song variation in island populations of the white-crowned sparrow (*Zonotrichia leucophrys nutalli*): innovation through recombination. *Behaviour*, 140, 947–963.

Slabbekoorn, H. & Peet, M. (2003). Birds sing at higher pitch in urban noise. *Nature*, 424, 267.

Slabbekoorn, H. & Smith, T. B. (2002). Bird song, ecology and speciation. *Proc. Roy. Soc. London B*, 357, 493–503.

Strigul, N. (2009). Can imitation explain dialect origins? *Ecol. Modelling*, 220, 2624–2639.

Williams, J. M. & Slater, P. J. B. (1990). Modelling bird song dialects: the influence of repertoire size and number of neighbours. *J. Theor. Biol.*, 145, 487–496.

Wright, T. F. (1996). Regional dialects in the contact call of a parrot. *Proc. Roy. Soc. London B*, 263, 867–872.

Yang, X.-J., Lei, F.-M., Wang, G., & Jesse, A. J. (2007). Syllable sharing and inter-individual syllable variation in Anna's hummingbird *Calypte anna* songs, in San Francisco, California. *Folia Zool.*, 56, 307–318.

Zann, R. (1997). Vocal learning in wild and domesticated zebra finches: signature cues for kin recognition or epiphenomena? In C. T. Snowdon & M. Hausberger (Eds.), *Social Influences on Vocal Development* (pp. 85–97).: Camridge: Cambridge University Press.

DEPARTMENT OF MATHEMATICS, VU UNIVERSITY AMSTERDAM, AMSTERDAM, THE NETHERLANDS
E-mail address: `r.planque@vu.nl`

DEPARTMENT OF MATHEMATICAL SCIENCES AND CENTRE FOR MATHEMATICAL BIOLOGY, UNIVERSITY OF BATH, BATH, UNITED KINGDOM

INSTITUTE OF BIOLOGY LEIDEN, LEIDEN UNIVERSITY, LEIDEN, THE NETHERLANDS