The Evolution of Dispersal in Random Environments and The Principle of Partial Control

Lee Altenberg
altenber@hawaii.edu

Abstract

McNamara and Dall (2011) identified novel relationships between the abundance of a species in different environments, the temporal properties of environmental change, and selection for or against dispersal. Here, the mathematics underlying these relationships in their two-environment model are investigated for arbitrary numbers of environments. The effect they described is quantified as the fitness-abundance covariance. The phase in the life cycle where the population is censused is crucial for the implications of the fitness-abundance covariance. These relationships are shown to connect to the population genetics literature on the Reduction Principle for the evolution of genetic systems and migration. Conditions that produce selection for increased unconditional dispersal are found to be new instances of departures from reduction described by the "Principle of Partial Control" proposed for the evolution of modifier genes. According to this principle, variation that only partially controls the processes that transform the transmitted information of organisms may be selected to increase these processes. Mathematical methods of Karlin, Friedland, and Elsner, Johnson, and Neumann, are central in generalizing the analysis.[1] Analysis of the adaptive landscape of the model shows that the evolution of conditional dispersal is very sensitive to the spectrum of genetic variation the population is capable of producing, and suggests that empirical study of particular species will require an evaluation of its variational properties.

1 Introduction

In analyzing a model of a population that disperses in a patchy environment subject to random environmental change, McNamara and Dall (2011) describe "how an underappreciated evolutionary process, which we term 'The Multiplier Effect', can limit the evolutionary value of responding adaptively to environmental cues, and thus favour the evolutionary persistence of otherwise paradoxical unconditional strategies." By "multiplier effect", McNamara and Dall mean,

---

[1]Dedicated to the memory of Professor Michael Neumann, one of whose many elegant theorems provides for a result presented here.

If a genotype is distributed in space and its fitness varies with location, then selection will change the spatial distribution of the genotype through its effect on population demography. This process can accumulate genotype members in locations to which they are well suited. This accumulation by selection is the multiplier effect.

It is possible, they discover, for the 'multiplier effect' to reverse — for there to be an excess of the population in the *worst* habitats — when there is very rapid environmental change. The environmental change they model is a Markov process where are large number of patches switch independently between two environments that produce different growth rates for a population of organisms. They find that for moderate rates of environmental change, populations will have higher asymptotic growth rates if they reduce their rate of unconditional dispersal between patches, which produces effective selection for lower dispersal.

Their key finding is that the reversal of the 'multiplier effect' due to rapid environmental change corresponds exactly with a reversal in the direction in which dispersal evolves: when abundance is greater on better habitats, lower dispersal is selected for; when abundance is greater on worse habitats because the environment changes so fast, there is selection for higher dispersal.

McNamara and Dall conclude their paper saying, "the multiplier effect may underpin the evolution and maintenance of unconditional strategies in many biological systems." This is indeed the case. Their results are in fact part of the phenomenon already known as the "reduction principle", which was first described as such in models for the evolution of linkage (Feldman, 1972), and subsequently extended to models for the evolution of mutation rates, gene conversion, dispersal, sexual reproduction, and even cultural transmission of traditionalism (Altenberg, 1984). The reduction principle also underlies other phenomena: the 'error catastrophe' in quasispecies dynamics, and the effect of population subdivision on the maintenance of genetic diversity.

The Reduction Principle can be stated, in a rather general form, as the widely exhibited phenomenon that *mixing reduces growth*, and *differential growth selects for reduced mixing.*

While the reduction phenomenon studied in McNamara and Dall (2011) is not a new concept, three particular aspects of their study are novel:

1. their discovery of conditions that cause mixing to *increase* growth — which addresses the open problem posed in Altenberg (2004, Open Question 3.1) as to the conditions that produce departures from the reduction principle;

2. that these departures from reduction emerge from very rapidly changing environments; and

3. that these departures from reduction correspond to reversals in the association between fitness and abundance in different environments.

McNamara and Dall produce these results from a two-environment model. A principal goal here is to generalize each of these findings to arbitrary numbers of environments. Insight on how to generalize them is provided by clues in their results. Some of these clues point to the main tool used to achieve the generalization, a theorem of the late Sam Karlin, to be described.

The property described by McNamara and Dall as 'the multiplier effect' is here made mathematically precise, as a positive covariance between *fitness* and the *excess of the stationary distribution* of the population above what it would be in the absence of differential growth rates, as censused just after dispersal. I refer to this quantity as the *fitness-abundance covariance*, which is a bit more descriptive and specific than the term 'multiplier effect', which already has long use as a concept in economics.

A critical aspect to use of the fitness-abundance covariance is the phase in the life cycle at which the census is taken. When McNamara and Dall say that "individuals are likely to find themselves in circumstances to which they are well-adapted," it matters where in its life cycle the individual finds itself — whether it is on its natal site or has already dispersed. McNamara and Dall do not explicitly address the phase at which they take their census, but their model shows it to be just after dispersal, before reproduction.

The issue of census phase is explicitly addressed here, and is shown to critically affect properties of the fitness-abundance covariance. For populations censused just after dispersal, one cannot say in general that "individuals are already likely to be on the better site." As a consequence of this phase dependence, a novel result found here is that by taking a census of the populations before and after reproduction, one can in certain situations infer a bound on the duration of changing environments.

A result in McNamara and Dall (2011) that garnered considerable attention is that " 'stupid strategies' could be best for the genes" (University of Exeter, 2011):

> One underappreciated consequence of the multiplier effect is that because individuals tend to be in locations to which they are well suited, its mere existence informs an organism that it is liable to be in favourable circumstances. This information can outweigh environmental cues to the contrary, so that an individual should place more weight on the fact it exists than on any additional cues of location quality. McNamara and Dall (2011)

The general analysis provided here produces results that seems to contradict the above: philopatry is never an evolutionarily stable strategy when there is any level of environmental change; it can always be invaded by organisms that disperse from the correct environments.

In an attempt to resolve the apparent contradiction, I take a closer examination of the adaptive landscape — the gradient of fitness over the space of conditional dispersal probabilities. What is found is that the evolutionarily stable state is highly sensitive to constraints on the organismal variability for dispersal probabilities. Slight changes in the constraints can shift the evolutionarily stable state from complete philopatry to complete dispersal from some environments. This sensitivity means that conditional dispersal may be a highly volatile trait evolutionarily. Moreover, to understand the evolution of any particular species requires an analysis of the constraints on the phenotype, and the probabilities of generating heritable variation in any phenotypic direction — in short, an evolvability analysis (Wagner and Altenberg, 1996).

While it is relatively straightforward to determine the long-term growth rates of different dispersal phenotypes, determining the likelihood that such phenotypes will be produced by the population plunges one into issues of the organism's perceptual and cognitive limits, ecological correlates, and the genotype-phenotype map, and requires specific empirical knowledge of the organism and its variability in order to address. This is perhaps why, as Levinton (1988, p. 494) insightfully writes, "Evolutionary biologists have been mainly concerned with the *fate of variability* in populations, not the *generation of variability*. . . . Whatever the reason, the time has come to reemphasize the study of the origin of variation." A principle finding here is that the evolutionary outcome is not determined by the adaptive landscape studied here, and we are pointed instead to examine the variational properties of each particular species in question.

1.1 The Reduction Principle and Fisher's Fundamental Theorem of Natural Selection

The intuition as to why there should be selection for lower dispersal in a population at a stationary balance between dispersal and selection is well expressed in the following explanation:

> Even in the absence of genetic variability for local adaptation in a spatially heterogeneous environment, migration will be selected against because on the average an individual will disperse to an environment worse than the one it was born in, since better environments harbor more individuals. (Olivieri et al., 1995).

This is a description of populations that have equilibrated to a balance between dispersal and differential growth. Fisher's Fundamental Theorem is that differential growth rates increase the mean fitness of the population by an amount equal to the variance in the growth rates. When the population is at a stationary distribution, however, this requires that dispersal *decrease* the mean fitness by exactly the same amount. Fisher uses the phrase "deterioration of the environment" (Fisher 1958; discussed in Price 1972) to describe this exact counterbalance to the variance in fitness that increases the mean fitness. But he includes mutation in this concept:

> . . . an equilibrium must be established in which the rate of elimination is equal to the rate of mutation. To put the matter in another way we may say that each mutation of this kind is allowed to contribute exactly as much to the genetic variance of fitness in the species as will provide a rate of improvement equivalent to the rate of deterioration caused by the continual occurrence of the mutation. (Fisher, 1958, p. 41)

Fisher was thinking of mutation, not dispersal, in the above. But as we shall see later, the same mathematics underlies both. Like a the mutation/selection balance Fisher describes, dispersal will generally be to lesser quality environments when the population has reached a growth/dispersal balance.

The interchangeability of many results in population genetics between mutation and dispersal reflects the fact that an organism's location, like its genotype, is transmissible information about its state, and its degree of preservation during transmission is itself an organismal phenotype and subject to evolution (Cavalli-Sforza and Feldman 1973; Karlin and McGregor 1974; Altenberg 1984, pp. 15–16, p. 178 Schauber et al. 2007; Odling-Smee 2007). The issue of the faithfulness of transmission brings us to the reduction principle.

2 A Review of the Reduction Principle

McNamara and Dall are more correct than perhaps even they realized in noting that their subject is an "under-appreciated evolutionary process". It is clear that awareness of the body of population genetics literature on the reduction principle has not fully percolated between disciplines. Karlin's (1982) key theorem on the reduction phenomenon, and its application to the evolution of dispersal Altenberg (1984), were independently duplicated recently by Kirkland et al. (2006). And McNamara and Dall (2011) were evidently unaware of the paper by Kirkland et al. (2006), published in a mathematics journal.

One main goal of this paper, therefore, is to provide a 'portal' to the reduction principle, its historical development, and methods of analysis for a broader audience. Here, I tie-in the work of McNamara and Dall (2011) to the larger stream of work on the reduction principle, and show that their work contributes toward answering one of the main open problems in the field: how departures from the reduction phenomenon are produced.

It may be appropriate to apologize for the density of equations in this paper, as equations nowadays are often being relegated to online-only supplements. But the subject of this paper is in fact mathematical methodology. It is the mathematics that creates a single conceptual and analytical framework for dispersal, recombination, mutation, random environments, and multiple genetic processes. To show how they all share in a single body of results requires we delve into the mathematics.

It should be noted that many theoretical studies constrain their analysis to models having only $n = 2$ patches or genotypes, to allow explicit calculation of the eigenvalues and eigenvectors (e.g. McNamara and Dall 2011, Steinmeyer and Wilke 2009). There are mathematical tools from the reduction principle literature, however — in particular the aforementioned theorems of Karlin — that make analytical results tractable for arbitrary n. Dissemination of these tools to a larger audience is another principal goal of this paper. They are laid out in *Methods*.

2.1 Development of the Reduction Principle

In the first analyses of genetic modifiers of mutation, recombination, and migration by Marc Feldman and coworkers in the 1970s, a common result kept appearing, which was that reduced levels of mutation, recombination, or migration would evolve when populations were near equilibrium under a balance between the forces of selection and transmission. The earliest appearance of the reduction phenomenon in the literature is perhaps Fisher's (1930, p. 130) assertion that "the presence of pairs of factors in the same chromosome, the selective advantage of each of which reverses that of the other, will always tend to diminish recombination, and therefore to increase the intensity of linkage in the chromosomes of that species." This claim was mathematically verified by Kimura (1956). Nei (1967; 1969) posed the first three-locus model for the evolution of recombination, with a modifier locus controlling the recombination between two loci under selection, and found that only reduced recombination would evolve. The first fixed-point stability analysis of modifiers of recombination

between two loci under viability selection was by Feldman (1972), who found that recombination would be reduced by evolution. Subsequent studies extended the reduction result to larger and larger spaces of models, including modifiers of:

dispersal: Karlin and McGregor (1972); Balkau and Feldman (1973); Karlin and McGregor (1974); Teague (1977); Asmussen (1983); Hastings (1983); Feldman and Liberman (1986); Liberman and Feldman (1989); Wiener and Feldman (1991, 1993);

recombination: Feldman (1972); Karlin and McGregor (1972); Feldman and Balkau (1972, 1973); Feldman and Krakauer (1976); Feldman et al. (1980); Liberman and Feldman (1986a); Feldman and Liberman (1986); and

mutation: Karlin and McGregor (1972); Liberman and Feldman (1986b); Feldman and Liberman (1986).

(Note that this literature prefers the term 'migration', while 'dispersal' is preferred in the ecology literature. Literature searches need to include both.)

These studies also extended the generality of the reduction results to include arbitrary large modified rates, arbitrary viability selection regimes, and multiple modifier alleles. They could only analyze the case of two patches or two alleles per selected locus, however, due to their use of closed-form solutions for the determinants or eigenvalues. Hastings (1983) is notable in extending the phenomenon to continuous spatial variation.

Feldman (1972) proposed that the essential direction of evolution for the recombination modifiers was reduction in the recombination rates. Shortly thereafter, Karlin and McGregor (1972, 1974) proposed an alternative idea, that the underlying governor for the direction of modifier evolution was the "Mean Fitness Principle". The Mean Fitness Principle proposed that a modifier allele increases when rare if and only if it changes the parameter it controls to a value that would increase the mean fitness of the population at equilibrium. Both reduction and mean fitness principles explained the known results at that time. However, Karlin and Carmelli (1975, Fig. 1) found an example where reducing recombination would *decrease* the mean fitness of the population, while Feldman et al. (1980) showed that, even for this example, an allele reducing recombination would grow in the population. Therefore, only the reduction principle remained unfalsified. Subsequent modifier gene studies have found other counterexamples to the mean fitness principle (Uyenoyama and Waller, 1991a,b; Wiener and Feldman, 1993). In Feldman et al. (1980) is where reduction was first referred to as a "principle".

2.2   Karlin's Theorems

During the time period of these developments, Karlin had, ironically, elucidated the mathematical foundations for the reduction principle himself — without realizing it.

Karlin was investigating a seemingly distant topic — how population subdivision would affect the maintenance of genetic variation. To understand how the protection of alleles against extinction depended on migration patterns and rates, Karlin (1976, 1982) developed two general theorems on the spectral radius of perturbations of migration-selection systems. The spectral radius is the growth rate for the whole group of genotypes that comprise the perturbation as they approach a stationary distribution among themselves.

These theorems show how, for two different kinds of variation in migration, a greater level of 'mixing' reduces the spectral radius of the stability matrix for the system, and thus may cause some alleles to lose their protection against extinction. Hence, greater levels of mixing would lead to fewer polymorphic alleles. Preparatory to this work was the paper by Friedland and Karlin (1975). The theorems first appear, without proof, in Karlin (1976, pp. 642–647), and with proof as Theorems 5.1 and 5.2 in Karlin (1982), restated as follows:

Theorem 1 (Karlin 1982, Theorem 5.1, pp. 114–116, 197–198). *Consider a family of stochastic matrices that commute and are symmetrizable to positive definite matrices:*

$$\mathcal{F} := \{\mathbf{M}_h = \mathbf{L}\mathbf{S}_h\mathbf{R} \colon \mathbf{M}_h\mathbf{M}_k = \mathbf{M}_k\mathbf{M}_h\}, \qquad (1)$$

where $\mathbf{L}$ and $\mathbf{R}$ are positive diagonal matrices, and each $\mathbf{S}_h$ is a positive definite symmetric real matrix. Let $\mathbf{D}$ be a positive diagonal matrix. Then for each $\mathbf{M}_h, \mathbf{M}_k \in \mathcal{F}$, the spectral radius, ρ, satisfies:

$$\rho(\mathbf{M}_h\mathbf{M}_k\mathbf{D}) \leq \rho(\mathbf{M}_k\mathbf{D}).$$

Theorem 2 (Karlin 1982, Theorem 5.2, pp. 117–118, 194–196). *Let $\mathbf{M}$ be a non-negative irreducible stochastic matrix. Consider the family of matrices*

$$\mathbf{M}(\alpha) = (1 - \alpha)\mathbf{I} + \alpha\mathbf{M}.$$

Then for any positive diagonal matrix $\mathbf{D}$, the spectral radius

$$\rho(\alpha) = \rho(\mathbf{M}(\alpha)\mathbf{D})$$

is decreasing as α increases (strictly provided $\mathbf{D} \neq d\mathbf{I}$).

In Theorem 5.1, 'more mixing' is produced the application of a second mixing operator; in Theorem 5.2, more mixing is produced by the equal scalar multiplication of all the transition probabilities between states. In both cases, greater mixing reduces the spectral radius, which represents the asymptotic growth rate of a rare allele in Karlin's analysis.

Theorems 5.1 and 5.2 display certain tradeoffs in generality. In Theorem 5.2, $\mathbf{M}$ may be any irreducible stochastic matrix, but the variation in the matrix family is restricted to a single parameter — the scaling of the transition probabilities. In Theorem 5.1 on the other hand, the variation in the matrix family is more general in that it has up to $n-1$ degrees of freedom to vary (see *Remark* for Lemma 22), but the matrix class itself is narrower with the constraint that they be symmetrizable.

Karlin's proof of Theorem 5.2 relied upon the recently minted variational formula for the spectral radius of Donsker and Varadhan (1975). These results on the reduction principle, and their means of generalization, all came into being in the same time period.

3 Application of Karlin's theorems to the Evolution of Dispersal and Genetic Systems

My own contribution to the reduction principle began with a conjecture by Marcus Feldman (1980, personal communication). The existence of polymorphisms for genes controlling recombination and mutation rates had been discovered theoretically by Feldman and Balkau (1973) and Feldman and Krakauer (1976)). Generalizing from these examples, Feldman conjectured that whenever a parameter controlled by a gene enters *linearly* into the recursion on the frequency dynamics, then a polymorphism for that gene would exist in which:

1. the population, when fixed on an allele producing a particular value of the linear parameter, is at an equilibrium;

2. each allele's average value of the parameter is equal to that particular value; and

3. the gene is in linkage equilibrium with the rest of the genome.

Because condition 2. was analogous to the condition for alleles under viability selection that their marginal fitnesses be equal at equilibrium, these polymorphisms were called 'viability-analogous, Hardy-Weinberg' (VAHW) modifier polymorphisms.

The repeated appearance of the VAHW polymorphisms, and the repeated occurrence of the reduction principle in models of different phenomena (recombination, mutation, and dispersal) prompted me to investigate the possible unification of these phenomena, which is provided in Altenberg (1984).

It turns out that the only way a parameter can enter linearly in the recursion is if it modifies transmission probabilities rather than fitnesses. The approach to unification was to represent all of the models in one general expression, in which the specifics of the transmission probabilities $P(i \leftarrow j, k)$ (parents j and k produce offspring i) are ignored, while the *variation* produced by the modifier locus is made explicit.

The form of variation studied was where the modifier gene produced an equal scaling, m, of all transmission probabilities between states, i.e. $mP(i \leftarrow j, k)$, when $j \neq i$ or $k \neq i$. The principle models that exhibited the reduction principle all incorporated this form of variation. Equal scaling of transmission probabilities occurs when a single transformative event acts on the transmitted information, and the modifier gene controls the rate of this event (Altenberg, 2011).

With this explicit representation of variation, the models that had exhibited the reduction principle had stability matrices of the form $\mathbf{M}(m)\mathbf{D}$ for newly introduced modifier alleles, where $\mathbf{M}(m) = (1-m)\mathbf{I} + m\mathbf{P}$ as in Karlin's theorem. Once this structure is made evident, application of Karlin's Theorem 5.2 immediately yields the result that the growth rate of a new modifier allele was a decreasing function of m, so if it reduced m below the current value in the population, it would invade, and if it increased m above the current level, it would go extinct.

Thus evolution would reduce the rates of all of these various processes, or others that had never been modeled before but which were covered by the general formulation. Prior studies needed to assume only two alleles under selection, or two patches subdividing the population, because they relied on closed-form solutions to determinants or eigenvalues. Karlin's theorem allowed the result to be generalized to arbitrary numbers of alleles and patches, arbitrary patterns of transformation, and arbitrary selection regimes.

It should be noted that modifiers of segregation distortion have altogether different dynamics that merit a separate classification (Altenberg, 1984, pp. 170–178).

Slight variation among different models led to separate treatments for modifiers of mutation and recombination (Altenberg 1984, pp. 106–169, Altenberg and Feldman 1987), modifiers of dispersal (Altenberg, 1984, pp. 77–81, 178–199), modifiers of rates of asexual vs. sexual reproduction (ibid. pp. 199–203), and culturally transmitted modifiers of cultural transmission — i.e. 'traditionalism' (ibid. pp. 203–206). All of these phenotypes manifest the reduction principle for the same underlying reason, the spectral radius property shown in Karlin's Theorem 5.2.

3.1 The Dispersal Modifier Results of Altenberg (1984)

The results on the evolution of dispersal modifiers in Altenberg (1984, pp. 77–81, 178–199) will be briefly reviewed, so that the work need not be duplicated, as has recently occurred (Kirkland et al., 2006).

The model is of an organism that has a multiple-stage life cycle, consisting of random mating, semelparous reproduction, selection on gametes, zygotes, and adults, and lastly, dispersal. The reproductive output of an organism depends on its patch and its diploid genotype for a gene under selection. The probability of dispersing between any two patches is scaled by a modifier gene. The model includes several generalizations of prior work:

- arbitrary numbers of patches;

- arbitrary dispersal patterns between patches, which may include cycles and asymmetry;

- dispersal of either adults or gametes (but not dispersal of zygotes, which breaks the Hardy-Weinberg frequencies of diploids and complicates the analysis);

- arbitrary hard or soft selection patterns on diploids and gametes;

- arbitrary numbers of alleles at a dispersal-modifying locus; and

- arbitrary number of alleles for the locus with patch-specific fitnesses.

Analysis is made of the evolutionary stability of populations near equilibrium. In order for any new modifier allele to grow or decline at a geometric rate, the equilibrium must possess variation in the reproductive rates among patches and/or genotypes. This variation was first identified as a property of equilibrium populations at mutation-selection balance by Haldane (1937), and was later called the 'genetic load' by Muller (1950).

The term "fitness load" was used in Altenberg (1984) to generalize the genetic load concept to circumstances where there may be no genes involved — in particular, to patches with different growth rates where the stationary distribution leaves some patches as sinks and others as sources, as they were later to be called (Pulliam, 1988). The term 'selection potential', $V := \max_i(D_i)/\mathrm{mean}(D_i) - 1$, was adopted in Altenberg and Feldman (1987) because of the analogy to potentials in physical systems, and because V was the actual maximum potential selective advantage that a modifier allele could accrue. $V > 0$ is necessary for any geometric growth in the modifier allele. The condition $V = 0$ corresponds to a population at an 'ideal free distribution' (Fretwell and Lucas, 1969; Fretwell, 1972).

For the dispersal modifier model in Altenberg (1984), a positive selection potential requires some differences at equilibrium among the terms

$$\frac{N^{\mathrm{S}}(e)\,\overline{w}(e,i)}{N^{\mathrm{D}}(e)\,\overline{w}(e)},\qquad(2)$$

over the environments e, and genotypes i, where

$N^{\mathrm{S}}(e)$ is the population size in environment e after selection, and $N^{\mathrm{D}}(e)$ after dispersal,

$\overline{w}(e)$ is the mean fitness in environment e,

$\overline{w}(e,i)$ is the mean fitness of the allele i under selection in environment e,

$N^{\mathrm{S}}(e) = N(e)\,\overline{w}(e)$ under hard selection, and $N^{\mathrm{S}}(e)$ is constant under soft selection.

One can see the two sources for a selection potential in (2): ecological, i.e. variation in $N^{\mathrm{S}}(e)/N^{\mathrm{D}}(e)$ (mentioned in the earlier quote of Olivieri et al. 1995), and genetic, i.e. variation in $\overline{w}(e,i)/\overline{w}(e)$.

Ideal free distributions having $V = 0$ may be produced by the "balanced mixture polymorphisms" discussed in Altenberg (1984, pp. 101–104, 129, 189–190, 218–222), which are synonymous with the Nash equilibria studied in Schreiber and Li (2011).

The main results obtained are the manifestation of the Reduction Principle for dispersal rates. First, we have this result for modifier allele with extreme effect:

Result. *3.27, Altenberg (1984, p. 195) A modifier allele which stops all migration will always increase when introduced to a population with an equilibrium selection potential, for any linkage to the locus under selection.*

For modifier alleles with intermediate effects on dispersal, tractability requires the assumption of tight linkage between the modifier locus and the selected locus. Under tight linkage, the stability matrix for the new modifier allele becomes a direct sum of blocks for each allele i under selection:

$$\boldsymbol{\epsilon}_i(t+1) = \mathbf{D}_1[(1-m)\mathbf{I} + m\bar{\mathbf{M}}]\mathbf{D}_2(i)\,\boldsymbol{\epsilon}_i(t)\qquad(3)$$

where $\bar{\mathbf{M}}$ is the matrix of average dispersal probabilities produced by modifier alleles in the equilibrium population, and

$$\mathbf{D}_1 = \mathbf{diag}\left[\frac{1}{N^{\mathrm{D}}(e)}\right]_{e=1}^{n_E},\ \mathbf{D}_2(i) = \mathbf{diag}\left[\frac{N^{\mathrm{S}}(e)\,\overline{w}(e,i)}{\overline{w}(e)}\right]_{e=1}^{n_E},$$

where n_E is the number of patches. Then the following is obtained:

Result. *3.28, Altenberg (1984, p. 199)*

1. *The new modifier allele, a, can change frequency at a geometric rate, that is, $\rho(\mathbf{M}_{\mathsf{a}}\mathbf{D}_1\mathbf{D}_2(i)) \neq 1$, only if there is an equilibrium selection potential in the population, so that $\mathbf{D}_1\mathbf{D}_2(i) \neq \mathbf{I}$.*

2. *The spectral radius for the new modifier allele, a, depends only on how its marginal migration matrix $\mathbf{M}_{\mathsf{a}}$ is related to the equilibrium marginal migration matrix $\bar{\mathbf{M}}$. The results of Theorem 3.14 for linear variation ... therefore apply directly:*

Theorem. *3.14, Altenberg (1984, p. 137): For a tightly linked modifier locus, when a new modifier allele,* a*, is introduced to a population at a stable viability-analogous, tensor product equilibrium (VAHW), where there is a variance in the marginal fitnesses of the selected types present, then for* m *as defined in* (3)*, the new modifier allele frequency will increase if* $m < 1$*, and it will be excluded if* $m > 1$*.*

Theorem 3.14 derives directly from Karlin's Theorem 5.2, which shows in addition that asymptotic growth rate of the new modifier allele increases as m decreases throughout the range of m.

Karlin's Theorem 5.2, and the dispersal modifier results above, have recently been duplicated by Kirkland et al. (2006, Theorem 3.1). They use a novel, structure-based proof for their version of Theorem 5.2, while Karlin used the Donsker and Varadhan formula for the spectral radius. They apply it to the evolution of unconditional dispersal, and prove a special case of Altenberg (1984, Result 3.28 and Theorem 3.14) where the genetics and life history stages are absent. Their results are extended to continuous time models by Schreiber and Lloyd-Smith (2009, online Appendix B), while Altenberg (2010) uses the Donsker and Varadhan formula to extend Theorem 5.2 to the continuous time case.

The results in Kirkland et al. (2006), while being special cases of Altenberg (1984) as far as the genetics are concerned, offer generalizations of the reduction principle in other new directions, namely, they generalize the work on density-dependent population regulation first addressed for dispersal modifiers by Asmussen (1983), and cover the general case where growth rates decrease with population size (Kirkland et al., 2006, Assumptions A1-A3). They also cover the case of reducible dispersal matrices (Theorem 4.4), the case of lossy dispersal (Assumption A4), and the fate of the modifier allele far from perturbation (Theorem 3.3). They examine conditional dispersers, and analyze the evolutionarily stable state in which dispersal has been conditioned to the point where the population reaches an ideal free distribution.

It should be noted that the ideal free distribution was proposed as ultimate evolutionarily stable state by Kimura (1967) in his 'principle of minimum genetic load'. Kimura was thinking about the evolution of mutation rates, not dispersal. But the driving force in each case — the genetic load for mutation, and the presence of sink and source populations for dispersal (Pulliam, 1988) — is mathematically the same phenomenon.

4   Departures from Reduction

While the reduction phenomenon occurs throughout a diverse class of evolutionary models, there are two principal classes in which departures from reduction are found. The

first class, which will not be further addressed here, comprises situations in which the population is continually kept far from equilibrium, due to genetic drift (e.g. the Hill-Robertson effect (Barton and Otto, 2005; Roze and Barton, 2006; Keightley and Otto, 2006), also Gillespie (1981a)), varying selection regimes (Charlesworth, 1976; Gillespie, 1981b; Ishii et al., 1989; Sasaki and Iwasa, 1987; Bergman and Feldman, 1990; Wiener and Tuljapurkar, 1994; Schreiber and Li, 2011; Blanquart and Gandon, 2010), or flux of beneficial mutations (Eshel, 1973a,b; Kessler and Levine, 1998).

The second class comprises cases of populations near equilibrium where multiple transformation processes act on the transmissible information of the organism. Studies of multiple transformation processes where departures from reduction are found include the evolution of:

- *recombination in the presence of mutation* (Feldman et al., 1980; Charlesworth, 1990; Otto and Feldman, 1997; Pylkov et al., 1998). The greatest attention has been given to this combination. The departures from the reduction result in this case are the basis of the 'deterministic mutation hypothesis' for the evolution of sex (Kondrashov, 1982, 1984; Kouyos et al., 2007).

- *recombination in the presence of dispersal* (Charlesworth and Charlesworth, 1979; Pylkov et al., 1998);

- *multiple mutation processes* (Altenberg, 1984, pp. 137–151);

- *recombination in the presence of segregation and syngamy* (which self-fertilization exposes in the recursion) (Charlesworth et al., 1979; Holsinger and Feldman, 1983a);

- *mutation in the presence of segregation and syngamy* (exposed in the recursion by self-fertilization (Holsinger and Feldman, 1983b), or fertility selection (Holsinger et al., 1986; Twomey and Feldman, 1990)).

It is notable that in their studies of dispersal in the presence of mutation, Wiener and Feldman (1991, 1993) found no departures from the reduction principle.

The pattern of departures from the reduction principle caused by multiple transformation processes was summarized in Altenberg (1984, pp. 149, 225–228) by a simple heuristic:

The principle of partial control: When the modifier gene has only partial control over the transformations occurring at loci under selection, then it may be possible for the part which it controls to evolve an increase in rates.

In several cases where multiple transformation processes produce departures from reduction, the stability matrix

on the modifier gene has the form

$$\mathbf{M}(m) = (1 - m)\mathbf{A} + m\mathbf{B}. \qquad (4)$$

Matrices of the form (4) also appear when the modifier gene is not tightly linked to the loci under selection (Feldman 1972, Altenberg 1984, p. 135, Altenberg and Feldman 1987, Altenberg 2009b). Karlin's Theorem 5.2 does not apply to such matrices, leaving an entire class of models as an unsolved open problem. In a survey of open problems in the spectral analysis of evolutionary dynamics (Altenberg, 2004), the following problem was posed:

Open Question (3.1 in Altenberg 2004). *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be irreducible stochastic matrices, and let* $\mathbf{D} \neq c\mathbf{I}$ *be a positive diagonal matrix. Define*

$$\mathbf{M}(\mu) = (1 - \mu)\mathbf{A} + \mu\mathbf{B}. \qquad (5)$$

For what conditions on $\mathbf{A}$*,* $\mathbf{B}$*, and* $\mathbf{D}$ *is the spectral radius* $\rho(\mathbf{M}(\mu)\mathbf{D})$ *strictly decreasing in* μ*, for* $0 \leq \mu \leq 1$*, or*

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\rho(\mathbf{M}(\mu)\mathbf{D}) < 0?$$

This open problem brings us back to the paper by McNamara and Dall (2011).

5   The Model of McNamara and Dall (2011)

The model of McNamara and Dall (2011, eq. (D.10)) is an example of 'partial control' (4), where we set $\mathbf{A} = \mathbf{P}$ and $\mathbf{B} = \boldsymbol{\pi}\,\mathbf{e}^{\top}$, $\boldsymbol{\pi}$ being the stationary distribution for stochastic matrix $\mathbf{P}$, i.e. $\mathbf{P}\boldsymbol{\pi} = \boldsymbol{\pi}$. A major point of interest is that McNamara and Dall find conditions on $\mathbf{P}$ that produce departures from the reduction phenomenon, providing another example of the principle of partial control, and contributing towards answering the open problem posed in Altenberg (2004), above. The recursion for their model is

$$\mathbf{z}(t + 1) = \mathbf{M}(m)\,\mathbf{D}\,\mathbf{z}(t), \qquad (6)$$

where

$$\mathbf{M}(m) := (1 - m)\mathbf{P} + m\,\boldsymbol{\pi}\,\mathbf{e}^{\top}. \qquad (7)$$

The control exerted by m over the transformations occurring in the system in (7) is only partial because the environment itself undergoes transformation, represented by $\mathbf{P}$, and the organism cannot eliminate $\mathbf{P}$, but only shift between $\mathbf{P}$ and $\boldsymbol{\pi}\,\mathbf{e}^{\top}$.

The McNamara and Dall model represents the following. Let $z_i(t)$ be the number of individuals in environment i at time t, and $z_i(t+1)$ be the number after one iteration of reproduction and dispersal.

1. An individual is born into a site with environment type i;

2. The individual reproduces on the site, and produces an average of D_i offspring when in environment i;

3. Each offspring disperses independently with probability m to a random site;

4. In one generation, sites of environment type j change randomly and independently to type i with probability M_{ij};

5. The sites have settled down to a stationary distribution, so the probability that the site will be in environment state i is π_i.

Recursion (6) in summation form is:

$$z_i(t + 1) = (1 - m)\sum_j P_{ij}D_j z_j(t) + m\pi_i\sum_j D_j z_j(t).$$

McNamara and Dall (2011) obtain analytical results for the case of $n = 2$ types of environment:

$$\begin{bmatrix} z_1(t + 1) \\ z_2(t + 1) \end{bmatrix} = \mathbf{MD}\begin{bmatrix} z_1(t) \\ z_2(t) \end{bmatrix}, \text{ where } \mathbf{D} = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix},$$

and

$$\mathbf{M} = (1-m)\begin{bmatrix} 1 - P_{21} & P_{12} \\ P_{21} & 1 - P_{21} \end{bmatrix} + m\begin{bmatrix} \pi_1 & \pi_1 \\ 1 - \pi_1 & 1 - \pi_1 \end{bmatrix}. \quad (8)$$

The model is notable for how it represents environmental randomness. The common way to model randomly changing environments would be to let $\mathbf{z}(t)$ represent the population size in each patch, $\mathbf{M}$ represent the dispersal between patches, and let the matrix of environment-specific growth rates, $\mathbf{D}$, be a random or time-dependent variable on each patch (e.g. Karlin 1982, pp. 90–92, 103–104, 140–145), yielding a system

$$\mathbf{z}(t) = \mathbf{M}\,\mathbf{D}_{(t)}\,\mathbf{M}\,\mathbf{D}_{(t-1)}\ldots\mathbf{M}\,\mathbf{D}_{(2)}\,\mathbf{M}\,\mathbf{D}_{(1)}\,\mathbf{z}(0). \qquad (9)$$

The analysis of such models can be challenging, requiring a resort to approximations and numerical analysis (see Gillespie 1981b, Tuljapurkar 1990, Wiener and Tuljapurkar 1994). Progress is being made in this area, however, for example the analysis of a two-cycle model for the evolution of dispersal by Schreiber and Li (2011).

When the random process of changing environments is independent among all the patches, as the number of patches becomes large, the system becomes deterministic in the same way that the Wright-Fisher model becomes deterministic for large populations. This allows one to stop keeping track of each *patch*, and just keep track of the number of individuals in each *environment* type, which is what McNamara and Dall do in (6). This tremendously simplifies the analysis.

5.1 Clues to the Generalization of the Results

The original motivation for this paper was to generalize the results of McNamara and Dall (2011) to an arbitrary number of environments, and to gain insight into why their model produces departures from the reduction phenomenon. Their results reveal four clues needed to solve this generalization:

1. *The harmonic mean*: McNamara and Dall (2011) find that departures from the reduction phenomenon are determined by the critical condition $\tau_1^{-1} + \tau_2^{-1} < 1$, where τ_i is the expected duration of environment i. This expression is part of the harmonic mean, $2/(\tau_1^{-1} + \tau_2^{-1})$. Could the harmonic mean of τ_i figure into a generalization of their results?

2. *The limiting distribution*: The two matrices in (7), $\mathbf{P}$, and $\boldsymbol{\pi}\mathbf{e}^\top$, are not arbitrary, but have the relation $\boldsymbol{\pi}\mathbf{e}^\top = \lim_{t \to \infty} \mathbf{P}^t$. This means, notably, that they commute: $\mathbf{P}(\boldsymbol{\pi}\mathbf{e}^\top) = (\boldsymbol{\pi}\mathbf{e}^\top)\mathbf{P} = \boldsymbol{\pi}\mathbf{e}^\top$, and thus satisfy one key condition of Karlin's Theorem 5.1.

3. *The second eigenvalue*: The terms τ_1^{-1} and τ_2^{-1} derive from the probabilities in $\mathbf{P}$: $\tau_1^{-1} = 1 - P_{11}$ and $\tau_2^{-1} = 1 - P_{22}$. The condition $\tau_1^{-1} + \tau_2^{-1} < 1$ translates to $1 < P_{11} + P_{22}$. Is it a coincidence that $P_{11} + P_{22}$ appears in the second eigenvalue of $\mathbf{P}$, $\lambda_2(\mathbf{P}) = 1 - P_{11} - P_{22}$? Because we see that the critical condition becomes $\lambda_2(\mathbf{P}) < 0$, which is precisely when $\mathbf{P}$ no longer meets the condition of Karlin's Theorem 5.1 that it be symmetrizable to a positive definite matrix. By extrapolation, if all the eigenvalues of $\mathbf{P}$ besides 1 are negative, could this be a general condition for departures from reduction?

4. *Symmetrizability*: Since clues 2. and 3. show the relationship between the results of McNamara and Dall an Karlin's Theorem 5.2, and we note that irreducible 2×2 matrices are always symmetrizable, might we want to retain symmetrizability in $\mathbf{P}$ as we try to generalize the results to $n \times n$ matrices?

By following the last clue and constraining $\mathbf{P}$ to be symmetrizable as in (1), we shall find it tractable to generalize the results of McNamara and Dall (2011), and we shall see that the conjectures prompted by the first and third clues are true.

Symmetrizable stochastic matrices are equivalent to the transition matrices of ergodic reversible Markov chains (Altenberg, 2011, Lemma 2). A Markov chain is reversible when the probability of cycles in one direction equals the probability of cycles in the opposite direction (Ross, 1983, Theorem 4.7.1, p. 127). In nature, directional cycles of environmental change may be more the rule than the exception, however. Whether cyclical environments would produce different results remains an open question.

We can step beyond the McNamara and Dall model and obtain a more general theorem for departures from reduction for the form $\mathbf{M}(m) = \mathbf{P}[(1-m)\mathbf{I} + m\mathbf{Q}]$, where $\mathbf{P}$ and $\mathbf{Q}$ satisfy (1). This is provided in Theorem 16 in *Results*. The theorem in Altenberg (2009a, 2011) that generalizes the reduction principle to the evolution of *mutation* rates among multiple loci turns out to be a special case of Theorem 16. This again illustrates the fact that genetic, spatial, cultural, and other transmissible information all belong to a single mathematical framework, and that results from one domain can often translate easily into results in other domains.

6 Results

McNamara and Dall (2011) describe their concept of a "multiplier effect" without ever giving it a precise mathematical definition. But it is clear from their usage in McNamara and Dall (2011, online Appendix A, Theorem A) that what they are thinking about can be summarized as the covariance between 1) the growth rates in each environment, and 2) the excess abundance of the population in that environment over what it would be without differential growth rates. This is defined explicitly below as *the fitness-abundance covariance*.

When organisms are semelparous, and generations are discrete and non-overlapping, there are two phases in the life cycle that one can census the population: before and after dispersal, or equivalently, after and before reproduction. Thus, the fitness-abundance covariance must be defined for both census phases.

The fitness-abundance covariance is an object of interest in its own right. Section 6.1 ventures beyond the specifics of the McNamara and Dall model and explores various properties of the fitness-abundance covariance for the completely general case of $\mathbf{z}(t+1) = \mathbf{MD}\mathbf{z}(t)$, where $\mathbf{M}$ is a stochastic matrix representing any process of change between states, and $\mathbf{D}$ represents the state-specific growth rates. The generality of results in Section 6.1 not only includes the McNamara and Dall model as a special case, but goes beyond models of dispersal since $\mathbf{M}$ can just as well represent a mutation matrix between genotypes whose fitnesses are D_i. The results can also apply to a rare genotype in a sexual population where $\mathbf{MD}$ represents the linear stability matrix on its growth.

Section 6.3 returns to the specific model of McNamara and Dall with the chief goal of generalizing the results to any number of environmental states. Here is where we pursue the clues described in the previous section.

For clarity, terminology and conventions are provided in Table 1.

6.1 The Fitness-Abundance Covariance

A precise definition needs to be given for the degree to which "individuals tend to be in locations to which they are well suited." While it may sound reasonable that an organism's "mere existence informs an organism that it is liable to be in favourable circumstances" (McNamara and Dall, 2011, p. 237), this is not generally true.

The following is an example where an organism is more likely to find itself in a sink habitat than a source habitat. The situation is where there is a small source patch within a large habitat of sink patches. We get a simple result if we assume the extremes: that sink habitats are lethal, and dispersing organisms recruit with fixed probabilities π_i to each patch, i, and the dispersal rate is m. Then $v_1 = 1 - m(1 - \pi_1)$ is the stationary proportion of the population in the source patch after dispersal. The stationary portion of the population in the source patch, v_1, can be made as small as one wishes by large dispersal rate m, and small π_1.

Clearly, *before* dispersal, all organisms in this example are in the source patch. So one must be clear about when in the life cycle one is speaking. An organism 'deciding' on whether to disperse or not is obviously at the pre-dispersal phase. But the post-dispersal phase is the phase that Mc-Namara and Dall use to measure the 'multiplier effect'.

Examination of the results of McNamara and Dall also reveals that when they speak of an organism being "liable to be in favourable circumstances," what they actually mean is that an organism is *more likely* to be in a favorable habitat than it *would be* if there were no growth advantage there, not that the organisms is actually *liable to be there*. This is the concept that I make precise as the *fitness-abundance covariance*. Even in this relative value of abundance, however, we will see that the fitness-abundance covariance is not always positive.

The fitness-abundance covariance relates three different sets of values: the environment-specific growth rates D_i, the stationary distribution in the *presence* of differential growth rates, referred to as v_i, and the stationary distribution in the *absence* of differential growth rates, referred to as π_i.

The stationary distribution for recursion $\mathbf{z}(t + 1) = \mathbf{M}\mathbf{D}\,\mathbf{z}(t)$ satisfies

$$\rho\,\mathbf{v} = \mathbf{M}\mathbf{D}\mathbf{v},$$

where $\mathbf{v}$ is the eigenvector of $\mathbf{M}\mathbf{D}$ associated with the largest eigenvalue of $\mathbf{M}\mathbf{D}$, ρ. This is called the right Perron vector. Throughout, $\mathbf{v}(\mathbf{A})$ will represent the right Perron vector of a matrix $\mathbf{A}$ (see Table 1).

The magnitude of ρ determines whether the population grows ($\rho > 1$) or declines ($\rho < 1$) or is stationary ($\rho = 1$), and ecological models typically impose some kind of negative density dependence so that as population size $\mathbf{z}$ gets large enough, ρ decreases with $\mathbf{z}$, and a stationary state of $\rho = 1$ can be attained. The problems addressed

---

Table 1: Definitions and Symbols

$\mathbf{A}, \mathbf{M}, \mathbf{D}, \mathbf{P}, \mathbf{S}$ or other boldface capital letters represent $n \times n$ matrices, and $\mathbf{v}, \mathbf{x}, \mathbf{y}, \mathbf{e}$, or other bold face lower case characters represent n-vectors; the identity matrix is $\mathbf{I}$; a scalar matrix is $c\,\mathbf{I}$ for $c \in \mathbb{R}$;

$A_{ij} \equiv [\mathbf{A}]_{ij}$ represents the elements of $\mathbf{A}$, $i, j = 1, \ldots, n$, and x_i represents the elements of $\mathbf{x}$;

$D_i \equiv [D]_{ii}$ represents the diagonal elements of diagonal matrix $\mathbf{D}$;

a positive diagonal matrix has $D_i > 0$, $i = 1, \ldots, n$;

$[\mathbf{A}]^i$ represents the ith row of matrix $\mathbf{A}$, and $[\mathbf{A}]_j$ represents the jth column.

$\mathbf{e}$ represents the unit vector, where all elements are 1;

$\mathbf{e}_j$ represents the jth basis vector, which has 1 at position j and 0 elsewhere;

$\mathbf{diag}[\mathbf{x}] \equiv \mathbf{D_x}$ is a diagonal matrix of the vector $\mathbf{x}$;

$\mathbf{A}^\top, \mathbf{z}^\top, \mathbf{e}^\top$, etc., represent the transpose;

$\lambda_i(\mathbf{A}) \equiv \lambda_{Ai}$, $i = 1 \ldots n$ represent the eigenvalues of $\mathbf{A}$;

symmetrizable to $\mathbf{S}$ means that an $n \times n$ matrix can be represented as a product $\mathbf{A} = \mathbf{L}\mathbf{S}\mathbf{R}$, where $\mathbf{S}$ is a symmetric real matrix, and $\mathbf{L}$ and $\mathbf{R}$ are positive diagonal matrices;

stochastic means an $n \times n$ matrix with nonnegative elements and whose columns (by convention here) sum to one (column stochastic);

positive definite means a matrix that is symmetric and has only positive eigenvalues;

irreducible means an $n \times n$ nonnegative matrix where for every i, j there is some t such that $[\mathbf{A}^t]_{ij} > 0$;

$\rho(\mathbf{A}) := \max_i |\lambda_i(\mathbf{A})|$ represents the spectral radius, the largest modulus of any eigenvalue of $\mathbf{A}$.

$\lambda_1(\mathbf{A})$ by convention will refer to the *Perron root* of a nonnegative irreducible matrix $\mathbf{A}$, which is the positive eigenvalue guaranteed by Perron-Frobenius theory (Seneta, 2006, Theorems 1.1, 1.5) to exist, to be the spectral radius, and to be as large as the modulus (i.e. magnitude) of any other eigenvalue. So $\lambda_1(\mathbf{A}) = \rho(\mathbf{A}) \geq |\lambda_i(\mathbf{A})|$ for $i = 2, \ldots, n$.

$\mathbf{v}(\mathbf{A})$ and $\mathbf{u}(\mathbf{A})^\top$ represent the right and left *Perron vectors* of nonnegative irreducible $\mathbf{A}$, the eigenvectors associated with the Perron root, guaranteed by Perron-Frobenius theory to be strictly positive. So $\mathbf{A}\,\mathbf{v}(\mathbf{A}) = \rho(\mathbf{A})\,\mathbf{v}(\mathbf{A})$, and $\mathbf{u}(\mathbf{A})^\top \mathbf{A} = \rho(\mathbf{A})\,\mathbf{u}(\mathbf{A})^\top$. By convention $\mathbf{e}^\top \mathbf{v}(\mathbf{A}) = 1$ and $\mathbf{u}(\mathbf{A})^\top \mathbf{v}(\mathbf{A}) = 1$.

$\mathbf{v} \equiv \mathbf{v}(\mathbf{A})$, $\mathbf{u} \equiv \mathbf{u}(\mathbf{A})$, and $\rho \equiv \rho(\mathbf{A})$, throughout, where $\mathbf{A}$ is obvious from context.

$\boldsymbol{\pi} \equiv \mathbf{v}(\mathbf{P})$ traditionally represents the stationary distribution of irreducible (column) stochastic matrix $\mathbf{P}$.

The harmonic mean of a set of numbers $\{\tau_i\}$ is

$$E_H(\tau_i) := \frac{1}{\dfrac{1}{n} \sum_{i=1}^{n} \dfrac{1}{\tau_i}} \ .$$

here do not concern the absolute value of ρ, but only the relative changes to ρ and $\mathbf{v}$ under changes in $\mathbf{M}$ and $\mathbf{D}$. For a general treatment of negative density dependence, Kirkland et al. (2006) provide a thorough analysis.

The stationary distribution in the presence if differential growth rates depends on the phase in the life cycle at which the census is taken. The life cycle consists of alternation between differential growth and dispersal, $\ldots \mathbf{DMDMDM} \ldots$. When censused just after dispersal, the stationary distribution is $\mathbf{v(MD)}$. Censused just before dispersal it is $\mathbf{v(DM)}$.

We see that $\mathbf{v(MD)}$ and $\mathbf{v(DM)}$ have a simple relationship from the cyclical structure. $\mathbf{D\,v(MD)}$ is the Perron vector of $\mathbf{DM}$ up to scaling, since

$$\mathbf{D[MD\,v(MD)]} = \rho(\mathbf{MD})\,\mathbf{D\,v(MD)}$$

When scaled to satisfy $\mathbf{e}^\top \mathbf{v(DM)} = 1$, one gets the relationship:

$$\mathbf{v(DM)} = \frac{1}{\rho(\mathbf{MD})}\mathbf{D\,v(MD)}. \qquad (10)$$

It should be noted that in continuous-time models such as quasispecies (Eigen and Schuster, 1977), selection and transformation happen simultaneously so there are no separate life cycle phases, hence no distinction between pre- and post-dispersal stationary states.

For semelparous organisms with discrete, non-overlapping generations, the fitness-abundance covariance is now defined for both phases of the life cycle.

Definition (Fitness-Abundance Covariance)**.**
   The fitness-abundance covariance is defined as the unweighted covariance between the environment-specific growth rates and the excess of the stationary distribution above the distribution that the population would attain in the absence of differential growth rates:

1. *Post-dispersal:*

$$\mathcal{FA}(\mathbf{MD}) := \mathrm{Cov}(D_i, v_i(\mathbf{MD}) - v_i(\mathbf{M}))$$
$$= \frac{1}{n}\sum_{i=1}^{n} D_i(v_i(\mathbf{MD}) - v_i(\mathbf{M}))$$
$$- \frac{1}{n}\sum_{i=1}^{n} D_i \; \frac{1}{n}\sum_{j=1}^{n}(v_j(\mathbf{MD}) - v_j(\mathbf{M})).$$

2. *Pre-dispersal:*

$$\mathcal{FA}(\mathbf{DM}) := \mathrm{Cov}(D_i, v_i(\mathbf{DM}) - v_i(\mathbf{M}))$$
$$= \frac{1}{n}\sum_{i=1}^{n} D_i(v_i(\mathbf{DM}) - v_i(\mathbf{M}))$$
$$- \frac{1}{n}\sum_{i=1}^{n} D_i \; \frac{1}{n}\sum_{i=1}^{n}(v_i(\mathbf{DM}) - v_i(\mathbf{M})).$$

Several elementary results are described. The first shows that the relationship between the pre- and post-dispersal fitness-abundance covariances is Fisher's Fundamental Theorem of Natural Selection in a slightly new context.

Theorem 3 (Fitness-Abundance Covariance and Census Phases)**.**
   Let $\mathbf{M}$ be an irreducible column stochastic matrix and $\mathbf{D}$ a positive diagonal matrix.
   Then

$$\mathcal{FA}(\mathbf{DM}) = \mathcal{FA}(\mathbf{MD}) + \frac{1}{n\,\rho(\mathbf{MD})}\mathrm{Var}_\mathbf{v}(D_i),$$

where $\mathrm{Var}_\mathbf{v}(D_i)$ is the $\mathbf{v(MD)}$-weighted variance of D_i,

$$\mathrm{Var}_\mathbf{v}(D_i) := \sum_{i=1}^{n} v_i(\mathbf{MD})D_i^2 - \left(\sum_{i=1}^{n} v_i(\mathbf{MD})D_i\right)^2.$$

Proof. Here, $\boldsymbol{\pi} \equiv \mathbf{v(M)}$, $\mathbf{v} \equiv \mathbf{v(MD)}$, and $\rho \equiv \rho(\mathbf{MD})$. We first note that

$$\sum_{i=1}^{n} D_i v_i = \mathbf{e}^\top \mathbf{D\,v} = \mathbf{e}^\top \mathbf{MD\,v} = \rho\,\mathbf{e}^\top \mathbf{v} = \rho.$$

$$\mathcal{FA}(\mathbf{MD}) = \mathrm{Cov}(D_i, v_i - \pi_i)$$
$$= \frac{1}{n}\sum_{i=1}^{n} D_i(v_i - \pi_i) - \frac{1}{n}\sum_{i=1}^{n} D_i \frac{1}{n}\sum_{j=1}^{n}(v_j - \pi_j)$$
$$= \frac{1}{n}\left(\rho - \sum_{i=1}^{n} D_i\pi_i\right) - 0. \qquad (11)$$

Substitution with (10) and (11) gives

$$\mathcal{FA}(\mathbf{DM}) = \mathrm{Cov}(D_i, v_i(\mathbf{DM}) - \pi_i)$$
$$= \frac{1}{n}\sum_{i=1}^{n} D_i(\frac{1}{\rho}D_i v_i - \pi_i) - \frac{1}{n^2}\sum_{i=1}^{n} D_i \sum_{j=1}^{n}(v_j(\mathbf{DM}) - \pi_j)$$
$$= \frac{1}{n}\left(\frac{1}{\rho}\sum_{i=1}^{n} D_i^2 v_i - \sum_{i=1}^{n} D_i\pi_i\right) - 0 \qquad (12)$$
$$= \frac{1}{n}\left(\frac{1}{\rho}\left[\sum_{i=1}^{n} D_i^2 v_i - (\sum_{i=1}^{n} D_i v_i)^2\right] + \rho - \sum_{i=1}^{n} D_i\pi_i\right) \qquad (13)$$
$$= \frac{1}{n\,\rho}\mathrm{Var}_\mathbf{v}(D_i) + \mathcal{FA}(\mathbf{MD}). \qquad \blacksquare$$

Corollary 4 (Derivatives of Fitness-Abundance Covariances and ρ)**.**
   Let $\mathbf{M}(m)$ be a family of irreducible stochastic matrices, differentiable in m, and assume $\mathbf{v}(\mathbf{M}(m)) = \boldsymbol{\pi}$ for all $m \in (0,1]$. Let $\mathbf{D}$ be a positive diagonal matrix. Set $\rho \equiv \rho(\mathbf{M}(m)\mathbf{D})$.
   Then

$$\frac{\mathrm{d}}{\mathrm{d}m}\mathcal{FA}(\mathbf{M}(m)\mathbf{D}) = \frac{1}{n}\frac{\mathrm{d}\rho}{\mathrm{d}m}, \qquad (14)$$

and

$$\frac{\mathrm{d}}{\mathrm{d}m}\mathcal{FA}(\mathbf{DM}(m))$$
$$=\frac{1}{n}\left(\frac{\mathrm{d}\rho}{\mathrm{d}m}+\frac{1}{\rho}\frac{\mathrm{d}}{\mathrm{d}m}\mathrm{Var}_{\mathbf{v}}(D_i)-\frac{1}{\rho^2}\mathrm{Var}_{\mathbf{v}}(D_i)\right). \quad (15)$$

Proof. Differentiation of (11) and (13) directly gives (14) and (15). ∎

By plain intuition we would expect the fitness-abundance covariance to be positive. But McNamara and Dall found circumstances in which the fitness-abundance covariance is negative just after dispersal. The question then remains, what about just before dispersal, when individuals are still in the environment whose growth rate they just replicated under? Here is where intuition suggests the fitness-abundance covariance should be positive. This is proven to be the case when $\mathbf{M}$ is the transition matrix of a reversible Markov chain with positive eigenvalues. But a counterexample is provided when $\mathbf{M}$ represents a periodic chain that cycles through the states, which has complex eigenvalues.

Theorem 5 (Positivity of the Pre-Dispersal Fitness-Abundance Covariance).

Let $\mathbf{M}$ be the transition matrix of an ergodic reversible Markov chain, with only nonnegative eigenvalues. Let $\mathbf{D} \neq c\mathbf{I}$ be a positive diagonal matrix.

Then

$$\mathcal{FA}(\mathbf{DM}) := \mathrm{Cov}(D_i, v_i(\mathbf{DM}) - v_i(\mathbf{M})) > 0.$$

Proof. Here, $\boldsymbol{\pi} \equiv \mathbf{v}(\mathbf{M})$, $\mathbf{v} \equiv \mathbf{v}(\mathbf{MD})$, and $\rho \equiv \rho(\mathbf{MD})$. From (12),

$$\mathcal{FA}(\mathbf{DM}) > 0 \iff \sum_{i=1}^{n} D_i^2 v_i > \rho \sum_{i=1}^{n} D_i \pi_i.$$

Since $\mathbf{D} \neq \mathbf{c}\,\mathbf{I}$, and $\mathbf{v} > \mathbf{0}$,

$$\mathrm{Var}_{\mathbf{v}}(D_i) = \sum_{i=1}^{n} D_i^2 v_i - (\sum_{i=1}^{n} D_i v_i)^2 > 0. \quad (16)$$

The condition that $\mathbf{M}$ be the transition matrix of an ergodic reversible Markov chain is equivalent to it being diagonally similar to a symmetric matrix (Keilson 1979, Proposition 1.3B; Altenberg 2011, Lemma 2). Since $\mathbf{M}$ has all nonnegative eigenvalues, that matrix is positive semidefinite. This allows application of the inequality in Friedland and Karlin (1975, Theorem 4.1): $\rho(\mathbf{DM}) \geq \sum_{i=1}^{n} D_i \pi_i$. In (16) this gives

$$\sum_{i=1}^{n} D_i^2 v_i > (\sum_{i=1}^{n} D_i v_i)^2 = \rho^2 \geq \rho \sum_{i=1}^{n} D_i \pi_i. \quad ∎$$

For a counterexample to the positivity of the pre-dispersal fitness-abundance covariance, we try a transition matrix $\mathbf{M}$ that is as far from Theorem 5 as possible, so the states are periodic and the eigenvalues other than 1 are complex roots of unity. This represents the situation of pelagic organisms along a gyre (e.g. Cowen et al. 2006).

Theorem 6. *Let $\mathbf{M}$ be an n-cyclic matrix,*

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ & & 1 & \ddots & \vdots & 0 \\ \vdots & & & \ddots & 0 & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Then for $\mathbf{D} \neq c\mathbf{I}$,

$$\mathcal{FA}(\mathbf{MD}) = \frac{1}{n}\left(\prod_{i=1}^{n} D_i^{1/n} - \frac{1}{n}\sum_{i=1}^{n} D_i\right) < 0 \quad (17)$$

Proof. For a matrix cyclic in this direction, given any $\mathbf{z} > 0$, $[\mathbf{MDz}]_{(i \bmod n)+1} = D_i z_i$. Thus $(\mathbf{MD})^n\mathbf{z} = (\prod_{i=1}^{n} D_i)\mathbf{z}$. So $(\mathbf{MD})^n\mathbf{v}(\mathbf{MD}) = (\prod_{i=1}^{n} D_i)\mathbf{v}(\mathbf{MD}) = \rho(\mathbf{MD})^n\mathbf{v}(\mathbf{MD})$. Hence $\rho(\mathbf{MD}) = \prod_{i=1}^{n} D_i^{1/n}$. Substitution in (11) gives (17). $\mathcal{FA}(\mathbf{MD})$ is negative because it is $1/n$ times the difference between the geometric and arithmetic means of D_i, which is always negative if not all D_i are equal (Steele, 2004, pp. 20–26). ∎

Theorem 7. *When the states are transformed in a cycle, it is possible for the pre-dispersal fitness-abundance covariance to be negative.*

Proof. An example is constructed. Let $\mathbf{M}$ represent the period-3 cycle of states $1 \to 2 \to 3 \to 1 \ldots$

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

and let $\mathbf{D} = \mathbf{diag}\left[D_1, D_2, D_3\right]$.

The spectral radius is $\rho = (D_1 D_2 D_3)^{1/3}$. By symmetry, $\pi_i = 1/3$, $i = 1, 2, 3$. Symbolic computation with $Mathematica^{TM}$ shows that

$\mathbf{v}(\mathbf{DM})$

$$= \begin{bmatrix} \left(1 + D_2^{2/3}/(D_1^{1/3} D_3^{1/3}) + (D_2^{1/3} D_3^{1/3})/D_1^{2/3}\right)^{-1} \\ \left(1 + (D_1^{1/3} D_3^{1/3})/D_2^{2/3} + D_3^{2/3}/(D_1^{1/3} D_2^{1/3})\right)^{-1} \\ \left(1 + (D_1^{1/3} D_2^{1/3})/D_3^{2/3} + D_1^{2/3}/(D_2^{1/3} D_3^{1/3})\right)^{-1} \end{bmatrix}.$$

A numerical survey shows that $\mathcal{FA}(\mathbf{DM})$ is positive over most values of (D_1, D_2, D_3) except for a very narrow range of $\mathbf{D}$ near the boundary where $\mathcal{FA}(\mathbf{DM})$

becomes negative. One such value is $(D_1, D_2, D_3) = (1/40000, 1, 1/8)$, which yields $\rho = 0.0146$, $\mathbf{v}(\mathbf{MD})^\top = (0.894, 0.0015, 0.105)$, $\mathbf{v}(\mathbf{DM})^\top = (0.0015, 0.105, 0.894)$, and $\mathcal{FA}(\mathbf{DM}) = -0.159$. ∎

The point of this odd counterexample is not that it represents something we might find in nature, but rather to say that we cannot entirely trust our intuition about the fitness-abundance covariance, and that something more subtle is going on mathematically than we might suppose.

6.2 Individual Stationary State Frequencies

Let us now examine the relationships between individual values of v_i, D_i and π_i.

McNamara and Dall show for the case of $n = 2$ that the post-dispersal fitness-abundance covariance is positive or negative depending on the durations of the environments, restated in their terms here:

Theorem 8 (McNamara and Dall (2011, online Appendix A, Theorem A)). *Let $n = 2$ in (7) to give (8).*

1. *If $\tau_1^{-1} + \tau_2^{-1} < 1$ then $\rho(\mathbf{MD}) > \sum_{i=1}^2 D_i \pi_i$ and*

   (a) $D_1 < D_2 \implies v_2(\mathbf{MD}) > v_2(\mathbf{M}) = \pi_2$

   (b) $D_1 > D_2 \implies v_2(\mathbf{MD}) < v_2(\mathbf{M}) = \pi_2$.

2. *If $\tau_1^{-1} + \tau_2^{-1} = 1$ then $v_i(\mathbf{MD}) = v_i(\mathbf{M}) = \pi_i$, $i = 1, 2$, and $\rho(\mathbf{MD}) = \sum_{i=1}^2 D_i \pi_i$.*

3. *If $\tau_1^{-1} + \tau_2^{-1} > 1$ then $\rho(\mathbf{MD}) < \sum_{i=1}^2 D_i \pi_i$ and*

   (a) $D_1 < D_2 \implies v_2(\mathbf{MD}) < v_2(\mathbf{M}) = \pi_2$

   (b) $D_1 > D_2 \implies v_2(\mathbf{MD}) > v_2(\mathbf{M}) = \pi_2$.

The third case exhibits the very counterintuitive behavior that increasing the reproductive output of an environment will *lower* the stationary proportion in that environment. We can compare this result to the following general theorem on how changes to a matrix affect its Perron vector:

Theorem 9 (Elsner et al. (1982, Theorem 2.1)).
Let $\mathbf{A}$ be an $n \times n$ nonnegative irreducible matrix. Then for any nonnegative n-vector $\mathbf{a} \geq\neq \mathbf{0}$, $i \neq j \in \{1, \ldots, n\}$,

$$\frac{v_i(\mathbf{A} + \mathbf{e}_i \mathbf{a}^\top)}{v_i(\mathbf{A})} > \frac{v_j(\mathbf{A} + \mathbf{e}_i \mathbf{a}^\top)}{v_j(\mathbf{A})}. \qquad (18)$$

It is more useful for us to put it in the following form:

Corollary 10 (Change in the Perron Vector).
When normalized to frequencies, $\mathbf{e}^\top \mathbf{v} = 1$, then $v_i(\mathbf{A} + \mathbf{e}_i \mathbf{a}^\top) > v_i(\mathbf{A})$.

Proof. The result follows immediately from rearrangement of (18) and summation using $\mathbf{e}^\top \mathbf{v} = 1$:

$$\sum_{j \neq i} \frac{v_j(\mathbf{A} + \mathbf{e}_i \mathbf{a}^\top)}{v_i(\mathbf{A} + \mathbf{e}_i \mathbf{a}^\top)} = \frac{1 - v_i(\mathbf{A} + \mathbf{e}_i \mathbf{a}^\top)}{v_i(\mathbf{A} + \mathbf{e}_i \mathbf{a}^\top)}$$

$$< \sum_{j \neq i} \frac{v_j(\mathbf{A})}{v_i(\mathbf{A})} = \frac{1 - v_i(\mathbf{A})}{v_i(\mathbf{A})}. \qquad ∎$$

In this case, the behavior of the Perron root follows our intuition that increasing the ith row of $\mathbf{A}$ should increase the stationary proportion of v_i.

Something must be very different, therefore, between theorems 8 and 9, since they both deal with changes in the Perron vector when elements of the matrix are changed. Theorem 8 produces counterintuitive results that depend on $\tau_1^{-1} + \tau_2^{-1}$, while Theorem 9 has no conditions on details of the matrix. How can this discrepancy be reconciled?

We must write $\mathbf{A}$ in terms of $\mathbf{M}$ and $\mathbf{D}$ to compare the two results. Let $[\mathbf{M}]^i$ be the ith row of $\mathbf{M}$. We can write

$$\mathbf{A} + \mathbf{e}_i \mathbf{a}^\top = \begin{bmatrix} D_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & D_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & D_i + \epsilon & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & D_n \end{bmatrix} \mathbf{M}$$

where $\mathbf{A} = \mathbf{DM}$ and $\mathbf{a} = \epsilon[\mathbf{M}]^i$, $\epsilon > 0$. Corollary 10 shows that increasing the reproductive output of environment i from D_i to $D_i + \epsilon$ increases the stationary proportion in environment i. In the limit $\epsilon \to 0$, this gives:

Corollary 11. *For irreducible column stochastic matrix $\mathbf{M}$ and positive diagonal matrix $\mathbf{D}$:*

$$\frac{\partial v_\kappa(\mathbf{DM})}{\partial D_\kappa} \geq 0. \qquad (19)$$

In the case $n = 2$, we have $D_2 > D_1 \iff D_2 = D_1 + \epsilon$, $\epsilon > 0$, $\mathbf{a}^\top = \epsilon(M_{21}, M_{22})$, and

$$\mathbf{A} + \mathbf{e}_2 \mathbf{a}^\top = \begin{bmatrix} D_1 & 0 \\ 0 & D_1 + \epsilon \end{bmatrix} \mathbf{M} = \mathbf{DM}.$$

So Theorem 9 gives $v_2(\mathbf{DM}) > v_2(\mathbf{M})$ regardless of any details of $\mathbf{M}$.

The discrepancy is resolved by noticing that the order of $\mathbf{M}$ and $\mathbf{D}$ is reversed between Theorem 8 and Theorem 9. The difference between the two is essentially in the phase of the life cycle at which the population is censused.

We can contrast (19) with the following:

Corollary 12. *For irreducible column stochastic matrix $\mathbf{M}$ and positive diagonal matrix $\mathbf{D}$,*

$$\frac{\partial}{\partial D_\kappa} \log(v_\kappa(\mathbf{MD})) \geq \frac{\partial}{\partial D_\kappa} \log\left(\frac{\rho(\mathbf{MD})}{D_\kappa}\right). \qquad (20)$$

Proof. Substitution of (10) in (19) and differentiation gives:

$$0 \leq \frac{\partial v_\kappa(\mathbf{DM})}{\partial D_\kappa} = \frac{\partial}{\partial D_\kappa}\left(\frac{D_\kappa v_\kappa(\mathbf{MD})}{\rho(\mathbf{MD})}\right)$$

$$= \frac{-1}{\rho^2}\frac{\partial\rho}{\partial D_\kappa}D_\kappa v_\kappa + \frac{1}{\rho}v_\kappa + \frac{1}{\rho}D_\kappa\frac{\partial v_\kappa}{\partial D_\kappa} \iff$$

$$0 \leq \frac{-1}{\rho}\frac{\partial\rho}{\partial D_\kappa} + \frac{1}{D_\kappa} + \frac{1}{v_\kappa}\frac{\partial v_\kappa}{\partial D_\kappa} \iff$$

$$\frac{\partial}{\partial D_\kappa}(\log(\rho) - \log(D_\kappa)) \leq \frac{\partial}{\partial D_\kappa}\log(v_\kappa). \blacksquare$$

Comparing (19) and (20) we see that the stationary distributions at different census phases behave differently.

To summarize:

- the portion in environment i censused before dispersal always increases with growth rate D_i;

- the portion in environment i censused after dispersal can, under the right environment transition matrix, *decrease* with increasing growth rate D_i.

This allows us to make inferences on the duration of the environments based on changes in the proportions of the population in each environment before and after reproduction:

Corollary 13 (Census Inference on Durations of Environments)**.**

Consider the model of McNamara and Dall, $\mathbf{z}(t+1) = \mathbf{M}(m)\mathbf{D}\mathbf{z}(t)$, with $\mathbf{M}(m) = [(1-m)\mathbf{P}+m\,\boldsymbol{\pi}\,\mathbf{e}^\top]$ (7), where $\boldsymbol{\pi} = \mathbf{P}\boldsymbol{\pi} = \mathbf{v}(\mathbf{P})$. At a stationary distribution, let $\mathbf{v} \equiv \mathbf{v}(\mathbf{MD})$ be the vector of proportions of individuals in each environment before reproduction, and $\mathbf{v}^R \equiv \mathbf{v}(\mathbf{DM})$ be the proportions after reproduction. For $n=2$ environments,

1. *If $v_1^R > \pi_1$, we know that $D_1 > D_2$.*

2. *If in addition, $v_1 < \pi_1$, then we know $E_H(\tau_1,\tau_2) < 2$; or if $v_1 > \pi_1$, then $E_H(\tau_1,\tau_2) > 2$.*

6.3 Generalization of The McNamara and Dall Model

We shift now from general $\mathbf{M}$ to the specific model of dispersal in randomly changing environments of McNamara and Dall (2011). First, we see how the direction of selection on unconditional dispersal corresponds to the sign of the post-dispersal fitness-abundance covariance.

Corollary 14 (McNamara and Dall Model with general n)**.**

Let $\mathbf{M}(m) := [(1-m)\mathbf{P}+m\,\boldsymbol{\pi}\,\mathbf{e}^\top]$, where $\mathbf{P}$ is an irreducible stochastic matrix, and $\mathbf{P}\boldsymbol{\pi} = \boldsymbol{\pi}$. Let $\mathbf{D} \neq c\,\mathbf{I}$ be a positive diagonal matrix. Set $\mathbf{v} \equiv \mathbf{v}(\mathbf{MD})$. Then

$$\mathcal{FA}(\mathbf{MD}) = \mathrm{Cov}(D_i, v_i - \pi_i) > 0$$
$$\iff \frac{\mathrm{d}}{\mathrm{d}m}\rho(\mathbf{M}(m)\mathbf{D}) < 0, \qquad (21)$$

(the reduction phenomenon) and

$$\mathcal{FA}(\mathbf{MD}) = \mathrm{Cov}(D_i, v_i - \pi_i) < 0 \qquad (22)$$
$$\iff \frac{\mathrm{d}}{\mathrm{d}m}\rho(\mathbf{M}(m)\mathbf{D}) > 0. \qquad (23)$$

(departure from reduction).

Proof. At $m = 1$,

$$\mathbf{M}(1)\mathbf{D}\,\mathbf{v} = \boldsymbol{\pi}\,\mathbf{e}^\top\mathbf{D}\,\mathbf{v} = \rho(\mathbf{M}(1)\mathbf{D})\,\boldsymbol{\pi},$$

so $\mathbf{v} = \boldsymbol{\pi}$, hence $\mathrm{Cov}(\mathbf{D}_i, v_i(\mathbf{M}(1)\mathbf{D})-\pi_i) = 0$. For $m < 1$, (21) and (23) follow from (14). $\blacksquare$

Remark. It should be noted that this correspondence between the reduction phenomenon and the sign of the post-dispersal fitness-abundance covariance is specific to McNamara and Dall's model, $\mathbf{M}(m) = (1-m)\mathbf{P} + m\mathbf{v}(\mathbf{P})\,\mathbf{e}^\top$. Shortly we will examine the more general $\mathbf{M}(m) = \mathbf{P}[(1-m)\mathbf{I} + m\mathbf{Q}]$, in which $\mathbf{v}(\mathbf{PQD}) \neq \mathbf{v}(\mathbf{P})$ generically, so $\mathrm{Cov}(\mathbf{D}_i, v_i(\mathbf{M}(1)\mathbf{D}) - v_i(\mathbf{P})) \neq 0$. Thus departures from reduction do not necessarily correspond to a negative fitness-abundance covariance. For the general open problem $\mathbf{M}(m) = (1-m)\mathbf{A} + m\mathbf{B}$ (5), it is not at all generic for $\mathbf{v}(\mathbf{AD}) = \mathbf{v}(\mathbf{A})$ or $\mathbf{v}(\mathbf{BD}) = \mathbf{v}(\mathbf{B})$, hence there is no general relationship between the reduction phenomenon and the sign of the fitness-abundance covariance.

Next, the expression in McNamara and Dall (2011) involving the durations of the environments, $\tau_1^{-1} + \tau_2^{-1}$, is generalized to the harmonic mean of the durations of n environments. The harmonic mean of the expected durations of states in a Markov chain (expected run lengths or 'exit times') is shown to have a fundamental relationship to the sum of the eigenvalues (the trace) of its transition matrix, an identity whose earliest reference I find is Shorrocks (1978, p. 1017), cited by Geweke et al. (1986):

Lemma 15 (Markov Chain Harmonic Mean and the Trace of the Transition Matrix)**.**

For a Markov chain with transition matrix $\mathbf{P}$, let τ_i be the expected duration of state i (the mean length of runs of i), and let $E_H(\tau_i)$ be their unweighted harmonic mean. Let $\lambda_i(\mathbf{P})$ be the eigenvalues of $\mathbf{P}$. These are related by the following:

$$E_H(\tau_i) := \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\tau_i}} = \frac{1}{1 - \frac{1}{n}\sum_{i=1}^{n}\lambda_i(\mathbf{P})} \geq 1, \qquad (24)$$

or, equivalently

$$E(\lambda_i(\mathbf{P})) := \frac{1}{n}\sum_{i=1}^{n}\lambda_i(\mathbf{P}) = 1 - \frac{1}{E_H(\tau_i)} \geq 0. \qquad (25)$$

Proof. The expected length of runs of any environmental state i is (Prais, 1955)

$$\tau_i := E(\text{duration of i}) = \sum_{t=0}^{\infty} t P_{ii}^{t-1} (1 - P_{ii})$$

$$= \sum_{t=0}^{\infty} (t+1) P_{ii}^{t} - \sum_{t=1}^{\infty} t P_{ii}^{t} = \sum_{t=0}^{\infty} P_{ii}^{t} = \frac{1}{1 - P_{ii}}.$$

Since $P_{ii} \geq 0$, then $\tau_i \geq 1$ so $E_H(\tau_i) \geq 1$. Since the trace of the matrix is $\sum_{i=1}^{n} P_{ii} = \sum_{i=1}^{n} \lambda_i(\mathbf{P})$, we have

$$\sum_{i=1}^{n} \lambda_i(\mathbf{P}) = \sum_{i=1}^{n} P_{ii} = \sum_{i=1}^{n} \left(1 - \frac{1}{\tau_i}\right) = n - \sum_{i=1}^{n} \frac{1}{\tau_i} \geq 0.$$

Rearrangements of the terms gives (24) and (25). ∎

The main result is now presented, tying together the eigenvalues of the environment transition matrix, $\mathbf{P}$, the effect of dispersal on the population growth rate, and the harmonic mean of environment durations. The result gives sufficient conditions in terms of the eigenvalues of $\mathbf{P}$ for departures from the reduction phenomenon.

This theorem must sacrifice some generality in the environment transition matrices in order to obtain tractability. The environmental change process must be a reversible Markov chain, which means it does not exhibit directional cycles (which requires complex eigenvalues in $\mathbf{P}$), which might be thought of as 'currents' through the set of states. This is the symmetrizability constraint that appears in Karlin's Theorem 5.1, and is required for technical reasons described in *Methods*. It remains an open problem whether the following theorem extends to all ergodic Markov chains.

Theorem 16 (Eigenvalues, Reduction Phenomenon, and Harmonic Mean of Environment Durations).

Let $\mathbf{P}$ and $\mathbf{Q} \in \mathbb{R}^{n,n}$ be transition matrices of reversible ergodic Markov chains that commute with each other. Let $\tau_i = 1/(1 - P_{ii})$ be the expected length of runs of state i under iteration of $\mathbf{P}$. Let

$$\mathbf{M}(m) := \mathbf{P}[(1 - m)\mathbf{I} + m\mathbf{Q}], \qquad (26)$$

and $\mathbf{D} \neq c\mathbf{I}$ be a positive diagonal matrix.

1. If all eigenvalues of $\mathbf{P}$ are positive, then

$$\frac{\mathrm{d}}{\mathrm{d}m} \rho(\mathbf{M}(m)\mathbf{D}) < 0, \qquad (27)$$

(the reduction phenomenon) and

$$E_H(\tau_i) > 1 + \frac{1}{n-1}. \qquad (28)$$

2. If all eigenvalues of $\mathbf{P}$ other than the Perron root 1 are negative, then

$$\frac{\mathrm{d}}{\mathrm{d}m} \rho(\mathbf{M}(m)\mathbf{D}) > 0, \qquad (29)$$

(departure from the reduction phenomenon) and

$$1 \leq E_H(\tau_i) < 1 + \frac{1}{n-1}. \qquad (30)$$

The proof is given in *Methods* section 7.1.

Remark. One should be careful here not to interpret this result as an implication *from* $E_H(\tau_i)$ to $\mathrm{d}\rho(\mathbf{M}(m)\mathbf{D})/\mathrm{d}m$. While it would be ideal to derive conditions for the reduction phenomenon from conditions on the durations of the environments, this is not possible here for $n \geq 3$; rather, both implications derive from the condition on the eigenvalues.

However, for $n = 2$ environments, the implication becomes possible, as seen in this slight generalization of McNamara and Dall (2011, Theorem B Corollary):

Corollary 17. *Let $\mathbf{P}$ and $\mathbf{Q}$ be 2×2 irreducible stochastic matrices that commute, and $\mathbf{D} \neq c\mathbf{I}$ be a positive diagonal 2×2 matrix. Then*

$$\frac{\mathrm{d}}{\mathrm{d}m} \rho(\mathbf{P}[(1 - m)\mathbf{I} + m\mathbf{Q}]\mathbf{D} > 0,$$

if and only if $P_{12} + P_{21} > 1$, or equivalently, $E_H(\tau_1, \tau_2) < 2$.

Proof. In the case of $n = 2$, there is only one other eigenvalue, $\lambda_2 = 1 - (P_{12} + P_{21})$. We have

$$E_H(\tau_1, \tau_2) = \frac{1}{1 - \frac{1}{2}[1 + 1 - (P_{12} + P_{21})]} = \frac{2}{P_{12} + P_{21}}.$$

So $P_{12} + P_{21} > 1 \iff E_H(\tau_1, \tau_2) < 2 \iff \lambda_2 < 0$, and by Theorem 16, $\dfrac{\mathrm{d}\rho(\mathbf{MD})}{\mathrm{d}m} > 0$. ∎

The model (8) of McNamara and Dall is a special case of Theorem 16 and Corollary 17. Note that all irreducible stochastic 2×2 matrices are transition matrices of ergodic reducible Markov chains:

Corollary 18. *Theorem 16 includes, as special cases, $\mathbf{Q} = \mathbf{P}^t$ for $t \geq 1$, and $\mathbf{Q} = \mathbf{P}^\infty = \boldsymbol{\pi}\mathbf{e}^\top$, where $\boldsymbol{\pi} = \mathbf{P}\boldsymbol{\pi}$ is the stationary distribution of $\mathbf{P}$.*

Proof. $\mathbf{P}$ and $\mathbf{P}^t$ commute, as do $\mathbf{P}$ and $\mathbf{P}^\infty = \boldsymbol{\pi}\mathbf{e}^\top$, since $\mathbf{P}\boldsymbol{\pi}\mathbf{e}^\top = \boldsymbol{\pi}\mathbf{e}^\top = \boldsymbol{\pi}\mathbf{e}^\top\mathbf{P}$. When $\mathbf{P}$ is the transition matrix of a reversible ergodic Markov chain, so too are $\mathbf{P}^t$ and $\mathbf{P}^\infty$. ∎

Theorem 16 is able to give us results only for the extrema of the distribution of eigenvalues of the environment

transition matrix, where all non-Perron eigenvalues are either positive or are all negative. It would be desirable to obtain some results for the interior region where there a mixture of positive and negative eigenvalues. Short of this, the location of the interior region can at least be made precise in terms of its relation to the space of matrices whose non-Perron eigenvalues are all negative.

This is done by forming a path from such extreme $\mathbf{P}$ to the opposite extreme, $\mathbf{I}$ — a homotopy $[0,1] \mapsto \{\mathbf{M}\} \subset \mathbb{R}^{n \times n}$. For $\mathbf{M}(m)$, the two endpoints of the path are $\mathbf{M}(0,m) = (1-m)\mathbf{I} + m\mathbf{Q}$, and $\mathbf{M}(1,m) = \mathbf{P}[(1-m)\mathbf{I} + m\mathbf{Q}]$. The path can be easily created by a convex combination, $\mathbf{M}(\alpha,m) = [(1-\alpha)\mathbf{I} + \alpha\mathbf{P}][(1-m)\mathbf{I} + m\mathbf{Q}]$, parameterized by $\alpha \in [0,1]$. It is straightforward from Lemma 15 that

$$E_H(\tau_i([1-\alpha]\mathbf{I} + \alpha\mathbf{P}) = \frac{1}{\alpha}E_H(\tau_i(\mathbf{P})),$$

so as α goes to 0, the harmonic mean of the environment durations goes to infinity. Using the homotopy between two extremes, we will now see that, for any $\mathbf{P}$, there is always some α below which the reduction phenomenon holds.

Corollary 19 (Convex Combination with an Extreme **P**)**.**

Let

$$\mathbf{M}(\alpha,m) := [(1-\alpha)\mathbf{I} + \alpha\mathbf{P}][(1-m)\mathbf{I} + m\mathbf{Q}],$$

where $\mathbf{P}$ and $\mathbf{Q}$ be transition matrices of ergodic reversible Markov chains that commute with each other, and $\alpha, m \in [0,1]$. Let $\mathbf{D} \neq c\mathbf{I}$ be a positive diagonal matrix.

Suppose that $\frac{\partial}{\partial m}\rho(\mathbf{M}(1,m)\mathbf{D}) > 0$ for $m \in (0,1]$. Then there exist critical values α_0, α_1 with $1/2 \leq \alpha_0 \leq \alpha_1 < 1$, such that for $m \in (0,1]$,

$$\frac{\partial}{\partial m}\rho(\mathbf{M}(\alpha,m)\mathbf{D}) < 0 \quad for \ \alpha \in [0,\alpha_0),$$

and

$$\frac{\partial}{\partial m}\rho(\mathbf{M}(\alpha,m)\mathbf{D}) > 0 \quad for \ \alpha \in (\alpha_1,1].$$

Proof. $\mathbf{M}(\alpha,0) = (1-\alpha)\mathbf{I} + \alpha\mathbf{P}$ plays the role of $\mathbf{P}$ in Theorem 16, so we need to know when $\lambda_i(\mathbf{M}(\alpha,0)) > 0 \ \forall i$. Let $\lambda_{\min}(\mathbf{P}) := \min_i \lambda_i(\mathbf{P})$. Then

$\lambda_{\min}([(1-\alpha)\mathbf{I} + \alpha\mathbf{P}]) = (1-\alpha) + \alpha\lambda_{\min}(\mathbf{P}) > 0 \iff$
$\lambda_{\min}(\mathbf{P}) > (\alpha-1)/\alpha = 1 - 1/\alpha.$

By Perron-Frobenius theory, irreducible stochastic $\mathbf{P}$ means $\lambda_{\min}(\mathbf{P}) > -1$. So $-1 > 1 - 1/\alpha$ (i.e. $\alpha \leq 1/2$) assures $\lambda_i(\mathbf{M}(\alpha,0)) > 0 \ \forall i$. Thus α_0 is no smaller than $1/2$.

Since $\rho(\mathbf{M}(\alpha,m)\mathbf{D})$ is a continuous function of α and m, we know $\partial\rho(\mathbf{M}(\alpha,m)\mathbf{D})/\partial m > 0$ for α in some neighborhood $(\alpha_1,1]$. ∎

So now we have characterized the interior regions as $[a_0, a_1]$ where the behavior of $\partial\rho(\mathbf{M}(\alpha,m)\mathbf{D})/\partial m$ needs to be characterized. We do not know, for example, if $\partial\rho/\partial m$ keeps the same sign for all m at a given α, or whether it can change sign more than once on $[\alpha_0, \alpha_1]$, and so forth. It remains an open problem.

Next, a particular kind of environmental change process is considered in which there is no causal connection between sequential environments. In other words, when the environment changes, it has no memory of its previous state. In genetics this is Kingman's 'House of Cards' model of mutation (Kingman, 1978, 1980)). It is shown for such a memoryless environment that the reduction phenomenon is the only possible outcome.

Theorem 20 ('House of Cards' Environmental Change)**.**
Let $\mathbf{M}(m)$ be defined as in (26) of Theorem 16. Let $\mathbf{D} \neq c\mathbf{I}$ be a positive diagonal matrix. Suppose that when the environment changes, its current state has no influence on its next state. Suppose further that the expected duration of an environment is $\tau_i = \tau$ for all environments i.

If $\tau = 1$, then $\dfrac{\mathrm{d}}{\mathrm{d}m}\rho(\mathbf{M}(m)\mathbf{D}) = 0$.

If $\tau > 1$, then $\dfrac{\mathrm{d}}{\mathrm{d}m}\rho(\mathbf{M}(m)\mathbf{D}) < 0$.

The proof is given in *Methods* section 7.2. Note that in this case, $\tau > 1$ becomes a sufficient condition for selection for reduced dispersal, which it is not in the general case in Theorem 16.

6.4   The Conditional Dispersal Model

The result of McNamara and Dall (2011) that drew particular attention was their finding that, under a broad range of circumstances, it is better for the organism to ignore cues about the environment and instead follow philopatry. The general form for their cue model is a modification of (7):

$$\mathbf{MD} = \mathbf{P}[(\mathbf{I} - \mathbf{C}) + \boldsymbol{\pi}\,\mathbf{e}^\top\mathbf{C}]\mathbf{D},$$

where $\mathbf{C}$ is a diagonal matrix of the conditional dispersal probabilities, C_i, that an individual disperses given it is in environment i.

A change in an organism's response to cues about its environment is manifest as changes to C_j, hence the object of interest is the change in asymptotic growth as the conditional dispersal rate is changed:

$$\frac{\partial}{\partial C_j}\rho(\mathbf{P}[(\mathbf{I} - \mathbf{C}) + \boldsymbol{\pi}\,\mathbf{e}^\top\mathbf{C}]\mathbf{D}).$$

The next result shows that an organism should increase its dispersal from any environment where its destinations correlate better with left Perron vector $\mathbf{u}(\mathbf{MD})$ than does staying put, and that in general there is *always* at least one such environment.

Theorem 21 (Conditional Dispersal).
Let $\mathbf{M} := \mathbf{P}[(\mathbf{I} - \mathbf{C}) + \boldsymbol{\pi}\,\mathbf{e}^{\top}\mathbf{C}]$, *where* $\mathbf{P}$ *be an irreducible stochastic matrix,* $\mathbf{C}$ *and* $\mathbf{D}$ *are positive diagonal matrices, with* $C_i \in (0,1)$, *and* $\boldsymbol{\pi} = \mathbf{P}\boldsymbol{\pi}$. *Refer to the left and right Perron vectors as* $\mathbf{u} \equiv \mathbf{u}(\mathbf{MD})$ *and* $\mathbf{v} \equiv \mathbf{v}(\mathbf{MD})$. *Then:*

1. *The derivative of the spectral radius with respect to each* C_{κ} *is*

$$\frac{\partial}{\partial C_{\kappa}}\rho(\mathbf{P}[(\mathbf{I}-\mathbf{C})+\boldsymbol{\pi}\,\mathbf{e}^{\top}\mathbf{C}]\mathbf{D})$$

$$= D_{\kappa}\,v_{\kappa}\sum_{i=1}^{n}u_i(\pi_i - P_{i\kappa}) \qquad (31)$$

$$= D_{\kappa}v_{\kappa}n\,[\mathrm{Cov}(u_i,\pi_i) - \mathrm{Cov}(u_i, P_{i\kappa})].$$

2. *There is always at least one* κ *for which*

$$\frac{\partial}{\partial C_{\kappa}}\rho(\mathbf{P}[(\mathbf{I}-\mathbf{C})+\boldsymbol{\pi}\,\mathbf{e}^{\top}\mathbf{C}]\mathbf{D}) > 0,$$

and at least one κ *for which*

$$\frac{\partial}{\partial C_{\kappa}}\rho(\mathbf{P}[(\mathbf{I}-\mathbf{C})+\boldsymbol{\pi}\,\mathbf{e}^{\top}\mathbf{C}]\mathbf{D}) < 0,$$

unless $\mathbf{D} = c\,\mathbf{I}$ *or* $\mathbf{P} = \boldsymbol{\pi}\,\mathbf{e}^{\top}$.

3. *The gradient of the spectral radius with respect to* $\mathbf{C}$ *is:*

$$\nabla_{\mathbf{C}}\,\rho(\mathbf{MD}) := \left[\frac{\partial\rho(\mathbf{MD})}{\partial C_{\kappa}}\right]_n^{\kappa=1} = \mathbf{u}^{\top}(\boldsymbol{\pi}\mathbf{v}^{\top} - \mathbf{P}\mathbf{D_v})\mathbf{D}.$$

4. *There is always a subspace,* $\mathcal{N}$, *of perturbations of* $\mathbf{C}$ *that are neutral for* $\rho(\mathbf{MD})$. $\mathcal{N} = \{\boldsymbol{\xi}\colon \nabla_{\mathbf{C}}\,\rho(\mathbf{MD})\,\boldsymbol{\xi} = 0\}$ *is an* $n-1$ *dimensional linear subspace. Its basis includes strictly positive* $\boldsymbol{\zeta} = \mathbf{D}_v^{-1}\mathbf{D}^{-1}\boldsymbol{\pi}$, *i.e.*

$$\zeta_{\kappa} = \frac{\pi_{\kappa}}{D_{\kappa}v_{\kappa}}. \qquad (32)$$

The proof is given in *Methods* section 7.3.

Theorem 21 shows that an organism can increase its asymptotic growth rate by dispersing more from any environment κ for which $\mathrm{Cov}(u_i(\mathbf{MD}),\pi_i) > \mathrm{Cov}(u_i(\mathbf{MD}), P_{i\kappa})$, which means that its environment if it disperses (distributed as $\boldsymbol{\pi}$), correlates better with $\mathbf{u}(\mathbf{MD})$ than its environment if it stays put (distributed as $[\mathbf{P}]_{\kappa}$).

Furthermore, there is always at least one such environment where *increased* dispersal is advantageous, and at least one other environment where *decreased* dispersal is advantageous, unless dispersal is neutral. Dispersal is neutral when growth rates are the same in all environments ($\mathbf{D} = c\,\mathbf{I}$), or the present environment has no influence on the next environment ($\mathbf{P} = \boldsymbol{\pi}\,\mathbf{e}^{\top}$).
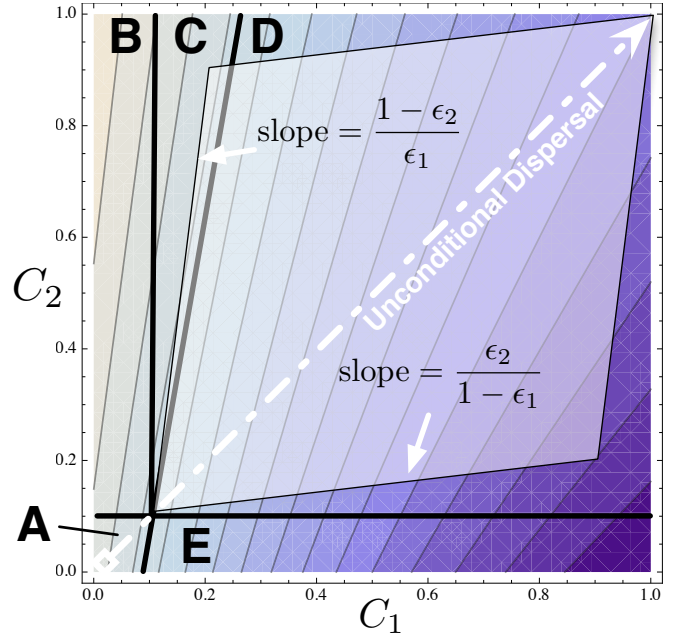

Figure 1:   Gradient of the asymptotic growth rate, $\rho(\mathbf{MD})$, over $(C_1, C_2) \in [0,1]^2$. Lighter means higher $\rho(\mathbf{MD})$. Model parameters: $(D_1, D_2) = (1.0, 0.5)$, $(p_{12}, p_{21}) = (0.204, 0.107)$. Diagonal dashed line $C_1 = C_2$ corresponds to variation in unconditional dispersal rates. Perturbations are away from $(C_1, C_2) = (0.1, 0.1)$. Regions A+B+C increase $\rho(\mathbf{MD})$. Regions D+E decrease $\rho(\mathbf{MD})$. McNamara and Dall constrain variation to fall within the parallelogram, with slope $\epsilon/(1-\epsilon)$ for the bottom, and $(1-\epsilon)/\epsilon$ for the side, where ϵ is the error rate for environmental cues; the error rate must be small enough for the parallelogram to enter region C for conditional dispersal to evolve. But the *unconstrained* ESS here is $(C_1, C_2) = (0, 1)$, and mutants anywhere in regions B+C increase conditional dispersal rate C_2 and are advantageous.

We see that Theorem 21 provides another situation that departs from the reduction phenomenon. Conditional dispersal is analogous to directed mutation (Cairns et al., 1988; Hall, 1990; Lenski and Mittler, 1993). I would not go so far as to say it exemplifies the principle of partial control, which was conceived for the situation of multiple *undirected* transformation processes, such as recombination in the presence of mutation. It is possible, however, to view conditional dispersal as control over only a part of the set of dispersal probabilities.

The existence of this departure from reduction holds for any environment transition matrix $\mathbf{P} \neq \boldsymbol{\pi}\,\mathbf{e}^{\top}$. The environment may even be constant, $\mathbf{P} = \mathbf{I}$, in which case as long as $\mathbf{D} \neq c\,\mathbf{I}$ so $\mathbf{u} \neq \mathbf{e}$, then $\frac{\partial}{\partial C_{\kappa}}\rho([(\mathbf{I} - \mathbf{C}) + \boldsymbol{\pi}\,\mathbf{e}^{\top}\mathbf{C}]\mathbf{D}) > 0$ for every κ where u_{κ} is below the $\boldsymbol{\pi}$-weighted average $\sum_{i=1}^{n}u_i\pi_i$ (see (31)).

Philopatry is clearly not the evolutionarily stable state (ESS) when there is environmental change, since there is

always at least one environment where conditional dispersal is advantageous. To understand how McNamara and Dall conclude that philopatry can be an ESS, it is helpful to look at the entire 'adaptive landscape' of $\rho(\mathbf{MD})$ as a function of conditional dispersal rates (C_1, C_2).

McNamara and Dall posit a species that can vary its probabilities, (p_1, p_2), of dispersing in response to a binary cue, where environment i produces the wrong cue with probability ϵ_i. The conditional dispersal rates are thus

$$\begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 1 - \epsilon_1 & \epsilon_1 \\ \epsilon_2 & 1 - \epsilon_2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} C_{\min} \\ C_{\min} \end{bmatrix}, \qquad (33)$$

Assuming that the species can vary (p_1, p_2) over the range $[0, 1] \times [0, 1]$, the variation in (C_1, C_2) falls within the parallelogram in Figure 1.

The example depicted in Figure 1 has a moderate rate of environmental change and differential growth between two environments. Darker means smaller $\rho(\mathbf{MD})$. Variation along the white diagonal line represents unconditional dispersal, and the gradient exhibits the reduction phenomenon. The contour lines of constant $\rho(\mathbf{MD})$ are shown by (32) to have slope

$$\left(\frac{\pi_2}{D_2 v_2} \right) \Big/ \left(\frac{\pi_1}{D_1 v_1} \right) = \frac{\pi_2 D_1 v_1}{\pi_1 D_2 v_2}.$$

The dispersal rates of the resident population are $(C_1, C_2) = (0.1, 0.1)$, the minimal dispersal attainable. The labeled regions A, B, C, D, and E demarcate the behavior where (C_1, C_2) departs from $(0.1, 0.1)$. Any mutant that falls within regions A, B, or C is advantageous. Regions B and C comprise increases in conditional dispersal from environment C_2. The slopes of the sides of the parallelogram derive from the cue error rate, ϵ.

Advantageous mutants arise only from in the intersection of the parallelogram and region C. For the intersection to be non-empty, ϵ must be small enough. When the error rate is so high that the parallelogram is contained entirely in region D, then the ESS is the lower left corner, the minimal value of dispersal.

We can see that the ESS is very sensitive to the error rate, however. With a slight decrease in ϵ, the ESS can shift from the lower left corner to the upper left corner of the parallelogram, which describes McNamara and Dall's result.

Therefore, in this adaptive landscape, it becomes very important how well the genetic and developmental system fills out the parallelogram with heritable variation. If the variation does not fully fill it out, at least two novel outcomes become possible. A slightly convex distribution overlapping region C would yield an intermediate level of dispersal as the fittest that occur. A slightly concave distribution would result in a bimodal distribution of the fittest phenotypes. This opens up potential for polymorphisms, disruptive selection, history dependence, or evolutionary volatility in the phenotype.

Moreover, the parallelogram is based on a particular model (33) of how organisms disperse. There is no categorical exclusion of genetic variation from accessing any point in the square $(C_1, C_2) \in [0, 1]^2$. Rather, it depends on the details of the organism, its capabilities, and its genotype-phenotype map. Any time an evolutionary outcome is sensitive to such details, one is bound to find interesting phenomena in the natural history.

7   Methods

The lengthier proofs for Theorems 16, 20, and 21 are now provided, prefaced by preparatory results, Lemma 22 and Theorem 23. It is here that we encounter the tractability afforded by using the transition matrices of reversible ergodic Markov chains, which is the key technique adopted from Friedland and Karlin (1975, Theorem 4.1) and Karlin's Theorem 5.1.

It should be noted that Karlin's Theorem 5.2 is not usable here, because of the presence of matrix $\mathbf{P}$ in $\mathbf{M}(m) = \mathbf{P}[(1-m)\mathbf{I}+m\mathbf{Q}]$. This is why it has been an open problem (Altenberg, 2004). An initial inroad on this open problem was obtained through application of elements from Karlin's Theorem 5.1 to the analysis of multivariate, multiple locus mutation rate evolution in Altenberg (2011). The application of these techniques is further extended here. Theorem 2 in Altenberg (2011) — a multivariate reduction principle for multiple loci in mutation-selection balance — is in fact a special case of (27) in Theorem 16 given here.

In the proofs to follow, (41), (43), (44) derive from Karlin's Theorem 5.1 proof. Other steps, including the use of a canonical form for symmetrizable $\mathbf{M}(m)$ (37), (38), (39), and (53) are drawn from the analysis in Altenberg (2011). Most of the remaining steps arise naturally from the problem, and may prove useful in other contexts.

Theorem 16 first requires a characterization of the spectral radius of $\mathbf{M}(m)\mathbf{D}$, which relies on the canonical form for $\mathbf{M}(m)$ that exists when it is constrained to be symmetrizable.

Lemma 22 (Canonical Form for Symmetrizable $\mathbf{M}(m)$).
Let $\mathbf{P}$ and $\mathbf{Q} \in \mathbb{R}^{n,n}$ be transition matrices of ergodic reversible Markov chains that commute with each other. Let

$$\mathbf{M}(m) := \mathbf{P}[(1 - m)\mathbf{I} + m\mathbf{Q}].$$

Then $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{M}$ can be decomposed as

$$\mathbf{P} = \mathbf{D}_{\boldsymbol{\pi}}^{1/2} \mathbf{K} \boldsymbol{\Lambda}_P \mathbf{K}^{\top} \mathbf{D}_{\boldsymbol{\pi}}^{-1/2}, \qquad (34)$$

$$\mathbf{Q} = \mathbf{D}_{\boldsymbol{\pi}}^{1/2} \mathbf{K} \boldsymbol{\Lambda}_Q \mathbf{K}^{\top} \mathbf{D}_{\boldsymbol{\pi}}^{-1/2}, \qquad (35)$$

and

$$\mathbf{M}(m) = \mathbf{D}_{\boldsymbol{\pi}}^{1/2} \mathbf{K} \boldsymbol{\Lambda}_P [(1 - m)\mathbf{I} + m\boldsymbol{\Lambda}_Q] \mathbf{K}^{\top} \mathbf{D}_{\boldsymbol{\pi}}^{-1/2}, \quad (36)$$

where $\mathbf{P}\boldsymbol{\pi} = \mathbf{Q}\boldsymbol{\pi} = \boldsymbol{\pi}$, with $\mathbf{e}^\top \boldsymbol{\pi} = 1$, $\mathbf{K}$ *is an orthogonal matrix, and* $\boldsymbol{\Lambda}_P$ *and* $\boldsymbol{\Lambda}_Q$ *are diagonal matrices of the eigenvalues of* $\mathbf{P}$ *and* $\mathbf{Q}$, *respectively.*

Furthermore, the first column of $\mathbf{K}$ *is* $[\mathbf{K}]_1 = \boldsymbol{\pi}^{1/2}$ *(element-wise square root).*

Proof. The transition matrices of ergodic reversible Markov chains, $\mathbf{M}$, can be represented in a canonical way (Keilson 1979, p. 33; Ababneh et al. 2006, p. 296; Altenberg 2011, Lemmas 1 and 2) as

$$\mathbf{M} = \mathbf{B}\mathbf{K}\boldsymbol{\Lambda}\mathbf{K}^\top\mathbf{B}^{-1}, \qquad (37)$$

where $\mathbf{B}$ is a positive diagonal matrix, unique up to scaling, and $\mathbf{K}$ is an orthogonal matrix, i.e. $\mathbf{K}\mathbf{K}^\top = \mathbf{K}^\top\mathbf{K} = \mathbf{I}$.

Any such $\mathbf{M}$ is clearly diagonalizable since $\boldsymbol{\Lambda}$ is a diagonal matrix. Diagonalizable $\mathbf{P}$ and $\mathbf{Q}$ commute by hypothesis, so they can be simultaneously diagonalized (Horn and Johnson, 1985, Theorem 1.3.19, p. 52), which means there exists invertible $\mathbf{X}$ such that

$$\mathbf{P} = \mathbf{X}\boldsymbol{\Lambda}_P\mathbf{X}^{-1} \text{ and } \mathbf{Q} = \mathbf{X}\boldsymbol{\Lambda}_Q\mathbf{X}^{-1}.$$

Hence $\mathbf{M}(m) = \mathbf{X}\boldsymbol{\Lambda}_P[(1-m)\mathbf{I} + m\boldsymbol{\Lambda}_Q]\mathbf{X}^{-1}$. Combining these two forms, the common matrix $\mathbf{X}$ can be represented as $\mathbf{X} = \mathbf{B}\mathbf{K}$.

Next, it is shown that

$$[\mathbf{K}]_1 = \boldsymbol{\pi}^{1/2} \qquad (38)$$

and

$$\mathbf{B} = c\,\mathbf{D}_{\boldsymbol{\pi}}^{1/2}, \ c > 0, \qquad (39)$$

satisfy $\mathbf{e}^\top\mathbf{P} = \mathbf{e}^\top$ and $\mathbf{P}\boldsymbol{\pi} = \boldsymbol{\pi}$. Since $\mathbf{K}$ is orthogonal, $[\mathbf{K}]_1 = \boldsymbol{\pi}^{1/2}$ if and only if $(\boldsymbol{\pi}^{1/2})^\top\mathbf{K} = \mathbf{e}_1^\top$, in which case, recalling that $\lambda_{P1} = 1$, substitution gives

$$\mathbf{e}^\top\mathbf{P} = \mathbf{e}^\top\mathbf{D}_{\boldsymbol{\pi}}^{1/2}\mathbf{K}\boldsymbol{\Lambda}_P\mathbf{K}^\top\mathbf{D}_{\boldsymbol{\pi}}^{-1/2} = \mathbf{e}_1^\top\boldsymbol{\Lambda}_P\mathbf{K}^\top\mathbf{D}_{\boldsymbol{\pi}}^{-1/2}$$
$$= \mathbf{e}_1^\top\mathbf{K}^\top\mathbf{D}_{\boldsymbol{\pi}}^{-1/2} = [\mathbf{K}]^1\mathbf{D}_{\boldsymbol{\pi}}^{-1/2} = \mathbf{e}^\top,$$

and

$$\mathbf{P}\boldsymbol{\pi} = \mathbf{D}_{\boldsymbol{\pi}}^{1/2}\mathbf{K}\boldsymbol{\Lambda}_P\mathbf{K}^\top\mathbf{D}_{\boldsymbol{\pi}}^{-1/2}\boldsymbol{\pi} = \mathbf{D}_{\boldsymbol{\pi}}^{1/2}\mathbf{K}\boldsymbol{\Lambda}_P\mathbf{K}^\top\boldsymbol{\pi}^{1/2}$$
$$= \mathbf{D}_{\boldsymbol{\pi}}^{1/2}\mathbf{K}\boldsymbol{\Lambda}_P\,\mathbf{e}_1 = \mathbf{D}_{\boldsymbol{\pi}}^{1/2}[\mathbf{K}]_1 = \boldsymbol{\pi}.$$

Substitution of $\mathbf{B} = c\,\mathbf{D}_{\boldsymbol{\pi}}^{1/2}$ in the form (37) produces (34), (35), and (36).

Remark. For any family of commuting symmetrizable stochastic matrices, $\mathbf{K}$ and $\mathbf{B}$ (up to scaling) are uniquely determined. Therefore, the only variation possible for the family is in λ_i, $i = 2, \ldots, n$, which means there are at most $n-1$ degrees of freedom of variation in the family. $\blacksquare$

Theorem 23 (The Spectral Radius).
Let $\mathbf{P}$ and $\mathbf{Q} \in \mathbb{R}^{n,n}$ *be transition matrices of ergodic reversible Markov chains that commute with each other, let* $\boldsymbol{\pi}$ *be their common right Perron vector, and let* $\{\lambda_{Pi}\}$ *and* $\{\lambda_{Qi}\}$ *be their eigenvalues. Let*

$$\mathbf{M}(m) := \mathbf{P}[(1-m)\mathbf{I} + m\mathbf{Q}].$$

Let $\mathbf{D}$ *be a positive diagonal matrix. Set* $\mathbf{v} \equiv \mathbf{v}(\mathbf{M}(m)\mathbf{D})$ *and* $\mathbf{u} \equiv \mathbf{u}(\mathbf{M}(m)\mathbf{D})$.
Then

$$\rho(\mathbf{M}(m)\mathbf{D}) = \sum_{i=1}^n \lambda_{Pi}[(1-m) + m\lambda_{Qi}]y_i^2, \qquad (40)$$

where

$$\mathbf{y} = (\mathbf{v}^\top\,\mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{D}\mathbf{v})^{-1/2}\,\mathbf{K}^\top\mathbf{D}_{\boldsymbol{\pi}}^{-1/2}\mathbf{D}\,\mathbf{v},$$

and $\mathbf{K}$ *is from the canonical form in Lemma 22.*

The left and right Perron vectors of $\mathbf{M}(m)\mathbf{D}$ *are related by*

$$\mathbf{u} = \frac{1}{(\mathbf{v}^\top\mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{D}\mathbf{v})}\mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{D}\,\mathbf{v}.$$

Proof. Canonical form (36) is used to produce a symmetric matrix similar to $\mathbf{M}(m)\mathbf{D}$, which allows use of the Rayleigh-Ritz formula for the spectral radius. The expression simplifies to a sum of terms involving the eigenvalues of the stochastic matrices $\mathbf{P}$ and $\mathbf{Q}$.

For brevity let $\boldsymbol{\Phi} := \mathbf{K}\boldsymbol{\Lambda}_P[(1-m)\mathbf{I} + m\boldsymbol{\Lambda}_Q]\mathbf{K}^\top$, so $\mathbf{M}(m) = \mathbf{B}\boldsymbol{\Phi}\mathbf{B}^{-1}$. Multiplication by $\mathbf{B}$, $\mathbf{D}^{1/2}$, and their inverses (where the positive diagonal $\mathbf{D}$ ensures the existence of $\mathbf{D}^{1/2}$ and $\mathbf{D}^{-1/2}$) gives the identities:

$$\rho(\mathbf{M}(m)\mathbf{D}) = \rho(\mathbf{B}\boldsymbol{\Phi}\mathbf{B}^{-1}\mathbf{D}) = \rho(\boldsymbol{\Phi}\mathbf{B}^{-1}\mathbf{D}\mathbf{B})$$
$$= \rho(\boldsymbol{\Phi}\mathbf{D}) = \rho(\mathbf{D}^{1/2}\boldsymbol{\Phi}\mathbf{D}^{1/2}) = \rho(\mathbf{S}),$$

where

$$\mathbf{S} := \mathbf{D}^{1/2}\boldsymbol{\Phi}\mathbf{D}^{1/2}$$
$$= \mathbf{D}^{1/2}\mathbf{K}\boldsymbol{\Lambda}_P[(1-m)\mathbf{I} + m\boldsymbol{\Lambda}_Q]\mathbf{K}^\top\mathbf{D}^{1/2}. \qquad (41)$$

Since $\mathbf{S}$ is symmetric, we may apply the Rayleigh-Ritz variational formula for the spectral radius (Horn and Johnson, 1985, Theorem 4.2.2, p. 176):

$$\rho(\mathbf{A}) = \max_{\mathbf{x}^\top\mathbf{x}=1} \mathbf{x}^\top\mathbf{A}\mathbf{x}. \qquad (42)$$

This yields

$$\rho(\mathbf{M}(m)\mathbf{D}) =$$
$$\max_{\mathbf{x}^\top\mathbf{x}=1} \mathbf{x}^\top\mathbf{D}^{1/2}\mathbf{K}\boldsymbol{\Lambda}_P[(1-m)\mathbf{I} + m\boldsymbol{\Lambda}_Q]\mathbf{K}^\top\mathbf{D}^{1/2}\mathbf{x}. \quad (43)$$

Since $\mathbf{M}$ is irreducible and $\mathbf{D}$ a positive diagonal matrix, $\mathbf{M}\mathbf{D}$ is irreducible, so by Perron-Frobenius theory there is a unique eigenvector $\hat{\mathbf{x}} > 0$ that yields the maximum in (43), allowing us to write

$$\rho(\mathbf{M}(m)\mathbf{D})$$
$$= \hat{\mathbf{x}}^\top\mathbf{D}^{1/2}\mathbf{K}\boldsymbol{\Lambda}_P[(1-m)\mathbf{I} + m\boldsymbol{\Lambda}_Q]\mathbf{K}^\top\mathbf{D}^{1/2}\hat{\mathbf{x}}. \qquad (44)$$

Define

$$\mathbf{y} := \mathbf{K}^\top \mathbf{D}^{1/2}\hat{\mathbf{x}}. \tag{45}$$

Substitution of (45) into (44) yields (40):

$$\begin{aligned}
\rho(\mathbf{M}(m)\mathbf{D}) &= \hat{\mathbf{x}}^\top \mathbf{D}^{1/2}\mathbf{K}\mathbf{\Lambda}_P[(1-m)\mathbf{I}+m\mathbf{\Lambda}_Q]\mathbf{K}^\top \mathbf{D}^{1/2}\hat{\mathbf{x}} \\
&= \mathbf{y}^\top \mathbf{\Lambda}_P[(1-m)\mathbf{I}+m\mathbf{\Lambda}_Q]\mathbf{y} \\
&= \sum_{i=1}^n \lambda_{Pi}[(1-m)+m\lambda_{Qi}]y_i^2.
\end{aligned}$$

Next, $\mathbf{y}$ will be solved in terms of $\mathbf{v}$ by solving for $\hat{\mathbf{x}}$, using the following two facts. For brevity, define $\mathbf{\Lambda}_{(m)} := \mathbf{\Lambda}_P[(1-m)\mathbf{I}+m\mathbf{\Lambda}_Q]$, and write $\mathbf{M} \equiv \mathbf{M}(m) = \mathbf{B}\mathbf{K}\mathbf{\Lambda}_{(m)}\mathbf{K}^\top \mathbf{B}^{-1}$:

1. $\rho(\mathbf{MD})\,\mathbf{v} = \mathbf{MDv} = \mathbf{B}\mathbf{K}\mathbf{\Lambda}_{(m)}\mathbf{K}^\top \mathbf{B}^{-1}\mathbf{Dv};$ $\qquad$ (46)

2. $\rho(\mathbf{MD})\,\hat{\mathbf{x}} = \mathbf{D}^{1/2}\mathbf{K}\mathbf{\Lambda}_{(m)}\mathbf{K}^\top \mathbf{D}^{1/2}\hat{\mathbf{x}}.$ $\qquad$ (47)

Multiplication on the left by $\mathbf{B}\mathbf{D}^{-1/2}$ in (47), and substitution of (46) reveals the right Perron vector of $\mathbf{MD}$:

$$\begin{aligned}
\rho(\mathbf{MD})\,(\mathbf{B}\mathbf{D}^{-1/2})\hat{\mathbf{x}} &= (\mathbf{B}\mathbf{D}^{-1/2})\mathbf{D}^{1/2}\mathbf{K}\mathbf{\Lambda}_{(m)}\mathbf{K}^\top \mathbf{D}^{1/2}\hat{\mathbf{x}} \\
&= \mathbf{B}\mathbf{K}\mathbf{\Lambda}_{(m)}\mathbf{K}^\top (\mathbf{B}^{-1}\mathbf{DBD}^{-1})\mathbf{D}^{1/2}\hat{\mathbf{x}}) \\
&= (\mathbf{B}\mathbf{K}\mathbf{\Lambda}_{(m)}\mathbf{K}^\top \mathbf{B}^{-1}\mathbf{D})(\mathbf{B}\mathbf{D}^{-1/2}\hat{\mathbf{x}}) \\
&= \mathbf{MD}(\mathbf{B}\mathbf{D}^{-1/2}\hat{\mathbf{x}}),
\end{aligned} \tag{48}$$

which shows that $\mathbf{B}\mathbf{D}^{-1/2}\hat{\mathbf{x}}$ is the right Perron vector of $\mathbf{MD}$, unique up to scaling, i.e.

$$\mathbf{v} = \mathbf{B}\mathbf{D}^{-1/2}\hat{\mathbf{x}} = \hat{c}\,\mathbf{D}_{\boldsymbol{\pi}}^{1/2}\mathbf{D}^{-1/2}\hat{\mathbf{x}},$$

for some $\hat{c}$ to be solved. This almost finishes the solution of $\hat{\mathbf{x}}$, giving

$$\hat{\mathbf{x}} = \frac{1}{\hat{c}}\mathbf{D}_{\boldsymbol{\pi}}^{-1/2}\mathbf{D}^{1/2}\,\mathbf{v}. \tag{49}$$

The constraint $\hat{\mathbf{x}}^\top \hat{\mathbf{x}} = 1$ gives

$$1 = \hat{\mathbf{x}}^\top \hat{\mathbf{x}} = \frac{1}{\hat{c}^2}(\mathbf{v}^\top \mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{D}\,\mathbf{v}),$$

so

$$\hat{c} = (\mathbf{v}^\top \mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{Dv})^{1/2}. \tag{50}$$

Substitution for $\hat{\mathbf{x}}$ now produces the expression in the theorem,

$$\begin{aligned}
\mathbf{y} &:= \mathbf{K}^\top \mathbf{D}^{1/2}\hat{\mathbf{x}} = \mathbf{K}^\top \mathbf{D}^{1/2}\frac{1}{\hat{c}}\mathbf{D}_{\boldsymbol{\pi}}^{-1/2}\mathbf{D}^{1/2}\,\mathbf{v} \\
&= (\mathbf{v}^\top \mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{Dv})^{-1/2}\,\mathbf{K}^\top \mathbf{D}_{\boldsymbol{\pi}}^{-1/2}\mathbf{D}\,\mathbf{v}.
\end{aligned} \tag{51}$$

By the same method as (48), $\mathbf{u}(\mathbf{M}(m)\mathbf{D})$ is derived, using multiplication on the right by $\mathbf{D}^{-1/2}\mathbf{B}$ to reveal the left Perron vector of $\mathbf{MD}$:

$$\begin{aligned}
\rho(\mathbf{MD})\,\hat{\mathbf{x}}^\top (\mathbf{D}^{1/2}\mathbf{B}^{-1}) &= \hat{\mathbf{x}}^\top \mathbf{D}^{1/2}\mathbf{K}\mathbf{\Lambda}_{(m)}\mathbf{K}^\top \mathbf{D}^{1/2}(\mathbf{D}^{1/2}\mathbf{B}^{-1}) \\
&= \hat{\mathbf{x}}^\top (\mathbf{D}^{1/2}\mathbf{B}^{-1})\mathbf{MD} = c^*\,\mathbf{u}(\mathbf{MD})^\top
\end{aligned}$$

for some $c^* > 0$. From $\mathbf{B} = c\,\mathbf{D}_{\boldsymbol{\pi}}^{1/2}$ (39), we get $(1/c^*)\hat{\mathbf{x}}^\top \mathbf{D}^{1/2}\mathbf{D}_{\boldsymbol{\pi}}^{-1/2} = \mathbf{u}(\mathbf{MD})^\top$. Substituting (49) and noting $\mathbf{u}^\top \mathbf{v} = 1$, we see the simple relationship to $\mathbf{v}(\mathbf{M}(m)\mathbf{D})$:

$$\mathbf{u}(\mathbf{M}(m)\mathbf{D}) = \frac{1}{(\mathbf{v}^\top \mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{Dv})}\mathbf{D}_{\boldsymbol{\pi}}^{-1}\mathbf{D}\,\mathbf{v}(\mathbf{M}(m)\mathbf{D}).$$

$\blacksquare$

One additional property that stems from the symmetrizability of $\mathbf{M}(m)$ in (26) is that $\rho(\mathbf{M}(m)\mathbf{D})$ is convex in m.

Theorem 24 (Convexity of $\rho(\mathbf{M}(m)\mathbf{D})$ in m)**.**
Let $\mathbf{P}$ and $\mathbf{Q} \in \mathbb{R}^{n,n}$ be transition matrices of ergodic reversible Markov chains that commute with each other. Let $\mathbf{D}$ be a positive diagonal matrix and

$$\mathbf{M}(m) := \mathbf{P}[(1-m)\mathbf{I}+m\mathbf{Q}].$$

Then $\rho(\mathbf{M}(m)\mathbf{D})$ is convex in m.

Proof. This follows the same lines as in Karlin (1982, Theorem F.1, p. 199). $\rho(\mathbf{M}(m)\mathbf{D}) = \rho(\mathbf{S})$ in (41), and $\mathbf{S} = (1-m)\mathbf{A} + m\mathbf{B}$, where $\mathbf{A} = \mathbf{D}^{1/2}\mathbf{K}\mathbf{\Lambda}_P\mathbf{K}^\top \mathbf{D}^{1/2}$ and $\mathbf{B} = \mathbf{D}^{1/2}\mathbf{K}\mathbf{\Lambda}_P\mathbf{\Lambda}_Q\mathbf{K}^\top \mathbf{D}^{1/2}$. The convexity of $\rho((1-m)\mathbf{A}+m\mathbf{B})$ is established by Lemma 25, to follow. $\blacksquare$

Lemma 25 (Convexity of the Spectral Radius)**.** *Let $\mathbf{A}$ and $\mathbf{B}$ be two symmetric matrices with unique eigenvectors $\hat{\mathbf{x}}_A$ and $\hat{\mathbf{x}}_B$ associated with their largest eigenvalue, normalized so $\hat{\mathbf{x}}_A^\top \hat{\mathbf{x}}_A = \hat{\mathbf{x}}_B^\top \hat{\mathbf{x}}_B = 1$.*
Then $\rho((1-m)\mathbf{A} + m\mathbf{B})$ is convex in m, and strictly convex if $\hat{\mathbf{x}}_A \neq \hat{\mathbf{x}}_B$

Proof. By hypothesis $\hat{\mathbf{x}}_A$ uniquely yields the maximum in (42), and likewise $\hat{\mathbf{x}}_B$ for $\mathbf{B}$, and $\hat{\mathbf{x}}_h$ for $(1-m)\mathbf{A} + m\mathbf{B}$. Therefore,

$$\begin{aligned}
\rho((1-m)\mathbf{A} + m\mathbf{B}) &= \hat{\mathbf{x}}_m^\top ((1-m)\mathbf{A}+m\mathbf{B})\hat{\mathbf{x}}_m \\
&= (1-m)\,\hat{\mathbf{x}}_m^\top \mathbf{A}\hat{\mathbf{x}}_m + m\,x_m^\top \mathbf{B}\hat{\mathbf{x}}_m \\
&\leq (1-m)\,\hat{\mathbf{x}}_A^\top \mathbf{A}\hat{\mathbf{x}}_A + m\,\hat{\mathbf{x}}_B^\top \mathbf{B}\hat{\mathbf{x}}_B = (1-m)\rho(\mathbf{A}) + m\rho(\mathbf{B}).
\end{aligned}$$

Equality requires $\hat{\mathbf{x}}_A = \hat{\mathbf{x}}_h = \hat{\mathbf{x}}_B$, because otherwise, $\hat{\mathbf{x}}_A^\top \mathbf{A}\hat{\mathbf{x}}_A > \hat{\mathbf{x}}_h^\top \mathbf{A}\hat{\mathbf{x}}_h$, or $\hat{\mathbf{x}}_B^\top \mathbf{B}\hat{\mathbf{x}}_B > \hat{\mathbf{x}}_h^\top \mathbf{B}\hat{\mathbf{x}}_h$, either of which produces strict inequality. $\blacksquare$

7.1 Proof of Theorem 16

Theorem 23 is now applied to the derivative of the spectral radius. The general relation is

$$\frac{\partial \rho(\mathbf{A})}{\partial m} = \mathbf{u}(\mathbf{A})^\top \frac{\partial \mathbf{A}}{\partial m}\mathbf{v}(\mathbf{A}) \tag{52}$$

(Caswell, 2000, Sec. 9.1.1).

This is derived for the specific case here by differentiating $\mathbf{S}\hat{\mathbf{x}} = \rho(\mathbf{MD})\hat{\mathbf{x}}$ (recall $\mathbf{S}$ from (41)), and then multiplying on the left by $\hat{\mathbf{x}}^{\top}$. Set $\rho \equiv \rho(\mathbf{MD})$.

$$\hat{\mathbf{x}}^{\top} \frac{\mathrm{d}(\mathbf{S}\hat{\mathbf{x}})}{\mathrm{d}m} = \hat{\mathbf{x}}^{\top} \left(\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}m}\hat{\mathbf{x}} + \mathbf{S}\frac{\mathrm{d}\hat{\mathbf{x}}}{\mathrm{d}m} \right) = \hat{\mathbf{x}}^{\top}\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}m}\hat{\mathbf{x}} + \rho\,\hat{\mathbf{x}}^{\top}\frac{\mathrm{d}\hat{\mathbf{x}}}{\mathrm{d}m}$$

$$= \hat{\mathbf{x}}^{\top}\frac{\mathrm{d}}{\mathrm{d}m}(\rho\hat{\mathbf{x}}) = \hat{\mathbf{x}}^{\top}\left(\frac{\mathrm{d}\rho}{\mathrm{d}m}\,\hat{\mathbf{x}} + \rho\,\frac{\mathrm{d}\hat{\mathbf{x}}}{\mathrm{d}m} \right) = \frac{\mathrm{d}\rho}{\mathrm{d}m} + \rho\,\hat{\mathbf{x}}^{\top}\frac{\mathrm{d}\hat{\mathbf{x}}}{\mathrm{d}m}.$$

Subtraction of $\rho(\mathbf{MD})\,\hat{\mathbf{x}}^{\top}\dfrac{\mathrm{d}\hat{\mathbf{x}}}{\mathrm{d}m}$ from both sides and substituting with (44) leaves:

$$\frac{\mathrm{d}\rho(\mathbf{MD})}{\mathrm{d}m} = \hat{\mathbf{x}}^{\top}\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}m}\hat{\mathbf{x}}$$

$$= \hat{\mathbf{x}}^{\top}\frac{\mathrm{d}}{\mathrm{d}m}\left[\mathbf{D}^{1/2}\mathbf{K}\mathbf{\Lambda}_P[(1-m)\mathbf{I} + m\mathbf{\Lambda}_Q]\mathbf{K}^{\top}\mathbf{D}^{1/2} \right]\hat{\mathbf{x}}$$

$$= \hat{\mathbf{x}}^{\top}\mathbf{D}^{1/2}\mathbf{K}\mathbf{\Lambda}_P[\mathbf{\Lambda}_Q - \mathbf{I}]\mathbf{K}^{\top}\mathbf{D}^{1/2}\hat{\mathbf{x}}.$$

Substitution with $\mathbf{y} := \mathbf{K}^{\top}\mathbf{D}^{1/2}\hat{\mathbf{x}}$ yields the derivative of (40):

$$\frac{\mathrm{d}\rho(\mathbf{MD})}{\mathrm{d}m} = \mathbf{y}^{\top}\mathbf{\Lambda}_P(\mathbf{\Lambda}_Q - \mathbf{I})\mathbf{y} = \sum_{i=1}^{n} \lambda_{Pi}(\lambda_{Qi} - 1)y_i^2. \quad (53)$$

Remark. Were $\mathbf{P}$ and $\mathbf{Q}$ not symmetrizable, but only diagonalizable and commuting, the analysis would arrive at an expression similar to (53) except that the nonnegative y_i^2 terms would be replaced by products whose signs we do not know, preventing further evaluation.

We know several things about the terms in the sum in (53):

1. Since $\mathbf{P}$ and $\mathbf{Q}$ are stochastic matrices, their Perron roots are 1, which here are labelled as $\lambda_{P1} = \lambda_{Q1} = 1$.

2. $\lambda_{Q1} - 1 = 0$. Thus the first term of the sum is zero.

3. $\lambda_{Qi} - 1 < 0$, for $i \in \{2, \ldots, n\}$, hence $(\lambda_{Qi} - 1)y_i^2 \leq 0$. Since $\mathbf{P}$ and $\mathbf{Q}$ are symmetrizable, $\lambda_{Pi}, \lambda_{Qi} \in \mathbb{R}$. Since $\mathbf{P}$ and $\mathbf{Q}$ are irreducible, by Perron-Frobenius theory Seneta (2006, Theorems 1.1, 1.5), eigenvalue 1 has multiplicity 1, and $|\lambda_{Qi}| \leq 1$, which together imply $\lambda_{Qi} < 1$ for $i \in \{2, \ldots, n\}$.

4. $y_i \neq 0$ for at least one $i \in \{2, \ldots, n\}$, whenever $\mathbf{D} \neq c\,\mathbf{I}$ for any $c > 0$. This fact will take a bit of work to show: Suppose to the contrary that $y_i = 0$ for all $i \in \{2, \ldots, n\}$. That means $\mathbf{y} = y_1\,\mathbf{e}_1$. Using $\hat{c}$ from (50), (51) becomes

$$y_1\,\mathbf{e}_1 = \hat{c}^{-1}\,\mathbf{K}^{\top}\mathbf{D}_{\boldsymbol{\pi}}^{-1/2}\mathbf{D}\,\mathbf{v}.$$

Multiplication on the left with $\mathbf{D}_{\boldsymbol{\pi}}^{1/2}\mathbf{K}$, and substitution with $[\mathbf{K}]_1 = \boldsymbol{\pi}^{1/2}$ (38) yields

$$y_1\,\mathbf{D}_{\boldsymbol{\pi}}^{1/2}\mathbf{K}\mathbf{e}_1 = y_1\,\mathbf{D}_{\boldsymbol{\pi}}^{1/2}[\mathbf{K}]_1 = y_1\,\boldsymbol{\pi} = \hat{c}^{-1}\mathbf{D}\,\mathbf{v}.$$

Multiplication of $y_1\hat{c}\,\boldsymbol{\pi} = \mathbf{D}\mathbf{v}$ by $\mathbf{M}$ gives

$$\mathbf{M}\boldsymbol{\pi}\,y_1\hat{c} = y_1\hat{c}\boldsymbol{\pi} = \mathbf{D}\mathbf{v} = \mathbf{M}\mathbf{D}\mathbf{v} = \rho(\mathbf{MD})\,\mathbf{v}.$$

Hence, $\mathbf{D}\mathbf{v} = \rho(\mathbf{MD})\,\mathbf{v}$, implying $\mathbf{D} = \rho(\mathbf{MD})\,\mathbf{I}$, contrary to hypothesis. Therefore, $\mathbf{D} \neq c\,\mathbf{I}$ for any $c > 0$ implies that $y_i \neq 0$ for at least one $i \in \{2, \ldots, n\}$.

Points 3., and 4. above together imply that $(\lambda_{Qi} - 1)y_i^2 < 0$ for at least one $i \in \{2, \ldots, n\}$. Inclusion of point 2. immediately implies for (53) that:

1. If $\lambda_{Pi} > 0$ for all i, then $\dfrac{\mathrm{d}}{\mathrm{d}m}\rho(\mathbf{MD}) < 0$.

2. If $\lambda_{Pi} < 0$ for $i = 2, \ldots, n$, then $\dfrac{\mathrm{d}}{\mathrm{d}m}\rho(\mathbf{MD}) > 0$.

3. Otherwise: there may be positive, negative, or zero terms $\lambda_{Pi}(\lambda_{Qi} - 1)y_i^2$, so the sign of $\dfrac{\mathrm{d}}{\mathrm{d}m}\rho(\mathbf{MD})$ depends on the particular values of the terms, of which we know little at this point.

Remark. Condition $\lambda_{Pi} > 0$ for all i is equivalent to $\mathbf{P}$ being symmetrizable to a positive definite matrix, which is the hypothesized condition in Karlin's Theorem 5.1. The condition $\lambda_{Pi} < 0$ for $i \in \{2, \ldots, n\}$ in case 2. happens to be the same as the well-known condition on the fitness matrix for a stable multiple-allele polymorphism (Kingman, 1961). Here this appears to be coincidence, rather than a clue to some deeper result. However, that condition is central to the 'viability-analogous' modifier polymorphisms, where the matrix $[1 - m_{ij}]_{i,j=1}^{n}$ (from diploid modifier genotypes $i|j$), must have all negative eigenvalues except the Perron root to assure stability of the modifier polymorphism (Feldman and Liberman, 1986; Liberman and Feldman, 1986a,b, 1989), supporting the analogy between viability coefficients and modifier values $1 - m_{ij}$.

The Harmonic Mean. The following inequalities are equivalent:

$$E_H(\tau_i) := \frac{1}{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\dfrac{1}{\tau_i}} < 1 + \frac{1}{n-1} = \frac{n}{n-1} \quad (54)$$

$$\iff n - 1 < \sum_{i=1}^{n}\frac{1}{\tau_i} = n - \sum_{i=1}^{n}P_{ii}$$

$$\iff 1 > \sum_{i=1}^{n}P_{ii} = \sum_{i=1}^{n}\lambda_i(\mathbf{P}) = 1 + \sum_{i=2}^{n}\lambda_i(\mathbf{P})$$

$$\iff 0 > \sum_{i=2}^{n}\lambda_i(\mathbf{P}). \quad (55)$$

Hence, $\lambda_i(\mathbf{P}) < 0$ for $i = 2, \ldots, n$ implies (55), or equivalently, (54). Conversely, $\lambda_i(\mathbf{P}) > 0$ for $i = 2, \ldots, n$ implies $\sum_{i=2}^{n}\lambda_i(\mathbf{P}) > 0$, hence (28). ∎

7.2 Proof of Theorem 20, 'House of Cards' Environmental Change

By hypothesis, the probability that environment remains unchanged in one generation is $\sigma = 1 - 1/\tau$. If the current environment has no influence on which environment comes next (Kingman's 'House of Cards' model, Kingman 1978, 1980) then

$$P_{ij} = (1 - \sigma)\pi_i + \sigma\delta_{ij}$$

where π_i be the probability that any changed environment becomes i, giving

$$\mathbf{P} = (1 - \sigma)\boldsymbol{\pi}\,\mathbf{e}^\top + \sigma\mathbf{I}.$$

Since $\boldsymbol{\pi}\,\mathbf{e}^\top$ is a rank-one matrix, $\lambda_i(\boldsymbol{\pi}\,\mathbf{e}^\top) = 0$ for $i = 2, \ldots, n$ (Horn and Johnson, 1985, p. 62). Hence $\lambda_i(\mathbf{P}) = \sigma$ for $i = 2, \ldots, n$.

For the case $\tau = 1$, then $\lambda_i(\mathbf{P}) = \sigma = 0$, $i = 2, \ldots, n$, so (53) evaluates to

$$\frac{\mathrm{d}\rho(\mathbf{MD})}{\mathrm{d}m} = \sum_{i=1}^{n} \lambda_{Pi}(\lambda_{Qi} - 1)y_i^2$$

$$= 1(1-1)y_1^2 + \sum_{i=2}^{n} 0(\lambda_{Qi} - 1)y_i^2 = 0.$$

For the case $\tau > 1$, then $\lambda_i(\mathbf{P}) = \sigma > 0$, $i = 2, \ldots, n$, so (53) evaluates to

$$\frac{\mathrm{d}\rho(\mathbf{MD})}{\mathrm{d}m} = \sum_{i=1}^{n} \lambda_{Pi}(\lambda_{Qi} - 1)y_i^2$$

$$= 1(1-1)y_1^2 + \sum_{i=2}^{n} \sigma(\lambda_{Qi} - 1)y_i^2 < 0,$$

since $\mathbf{D} \neq c\mathbf{I}$ for any $c \in \mathbb{R}$ implies that $\lambda_{Qi} - 1 < 0$ for $i \in \{2, \ldots, n\}$, and by 4. in the proof of Theorem 16, $y_i \neq 0$ for some for $i \in \{2, \ldots, n\}$. ∎

7.3 Proof of Theorem 21, Conditional Dispersal.

1. Basic identities used are $\partial\mathbf{C}/\partial C_\kappa = \mathbf{D}_{\mathbf{e}_\kappa}$, $\mathbf{e}^\top\mathbf{e}_\kappa = 1$, and $\mathbf{Pe}_\kappa = [\mathbf{P}]_\kappa$. The derivative formula (52), $\partial\rho(\mathbf{A})/\partial\beta = \mathbf{u}(\mathbf{A})^\top(\partial\mathbf{A}/\partial\beta)\mathbf{v}(\mathbf{A})$ (Caswell, 2000, Sec. 9.1.1), with respect to parameter β, is applied to yield

$$\frac{\partial}{\partial C_\kappa}\rho(\mathbf{MD}) = \frac{\partial}{\partial C_\kappa}\rho(\mathbf{P}[(\mathbf{I} - \mathbf{C}) + \boldsymbol{\pi}\,\mathbf{e}^\top\mathbf{C}]\mathbf{D})$$

$$= \mathbf{u}^\top\frac{\partial\mathbf{M}}{\partial C_\kappa}\mathbf{Dv} = \mathbf{u}^\top\mathbf{P}(-\mathbf{D}_{\mathbf{e}_\kappa} + \boldsymbol{\pi}\,\mathbf{e}^\top\mathbf{D}_{\mathbf{e}_\kappa})\mathbf{Dv}$$

$$= \mathbf{u}^\top(\mathbf{P}\boldsymbol{\pi}\,\mathbf{e}^\top\mathbf{e}_\kappa - \mathbf{Pe}_\kappa)\,D_\kappa\,v_\kappa \tag{56}$$

$$= \mathbf{u}^\top(\boldsymbol{\pi} - [\mathbf{P}]_\kappa)D_\kappa\,v_\kappa = D_\kappa\,v_\kappa\sum_{i=1}^{n} u_i(\pi_i - P_{i\kappa})$$

$$= D_\kappa\,v_\kappa\,n\,[\mathrm{Cov}(u_i, \pi_i) - \mathrm{Cov}(u_i, P_{i\kappa})]. \tag{57}$$

Note: $1/n^2$ in the covariance terms cancels in (57), e.g.

$$\mathrm{Cov}(u_i, P_{i\kappa}) = \frac{1}{n}\sum_i u_i P_{i\kappa} - \frac{1}{n^2}\sum_i u_i\sum_i P_{i\kappa}$$

$$= \frac{1}{n}\sum_i u_i P_{i\kappa} - \frac{1}{n^2}.$$

2. There is always at least one environment in which increased dispersal is advantageous, and at least one environment where decreased dispersal is advantageous: If no environment selects for increased dispersal, that means $\partial\rho(\mathbf{MD})/\partial C_\kappa \leq 0$ for all κ, hence $\mathbf{u}^\top(\boldsymbol{\pi} - [\mathbf{P}]_\kappa) \leq 0$, or, combined, $\mathbf{u}^\top(\boldsymbol{\pi}\,\mathbf{e}^\top - \mathbf{P}) \leq \mathbf{0}^\top$. Then $\mathbf{u}^\top(\boldsymbol{\pi}\,\mathbf{e}^\top - \mathbf{P})\boldsymbol{\pi} \leq 0$. But

$$\mathbf{u}^\top(\boldsymbol{\pi}\,\mathbf{e}^\top - \mathbf{P})\boldsymbol{\pi} = \mathbf{u}^\top\boldsymbol{\pi} - \mathbf{u}^\top\boldsymbol{\pi} = 0,$$

so $\mathbf{u}^\top(\boldsymbol{\pi} - [\mathbf{P}]_\kappa) = 0$ for all κ, which implies either $\boldsymbol{\pi} = [\mathbf{P}]_\kappa \ \forall \ \kappa$ since $\mathbf{u} > \mathbf{0}$, or $\mathbf{u}^\top = \mathbf{e}^\top$ which requires $\mathbf{D} = c\mathbf{I}$ for some $c \in \mathbb{R}$. If neither $\boldsymbol{\pi} = [\mathbf{P}]_\kappa \ \forall \ \kappa$ nor $\mathbf{D} = c\mathbf{I}$, then there must be some κ for which $\mathbf{u}^\top(\boldsymbol{\pi} - [\mathbf{P}]_\kappa) > 0$, hence $\frac{\partial}{\partial C_\kappa}\rho(\mathbf{P}[(\mathbf{I} - \mathbf{C}) + \boldsymbol{\pi}\,\mathbf{e}^\top\mathbf{C}]\mathbf{D}) > 0$. The parallel argument follows when $\leq$ is replaced by $\geq$ above.

   Remark. When $n = 2$, then it must be the case that the spectral radius is maximized at either $(C_1, C_2) = (1, 0)$, or at $(C_1, C_2) = (0, 1)$. This is illustrated in the numerical example in Figure 1.

3. A row vector is made from (56), over κ:

$$\nabla_\mathbf{C}\,\rho(\mathbf{MD}) := \left[\frac{\partial\rho(\mathbf{MD})}{\partial C_\kappa}\right]_n^{\kappa=1} = \left[\mathbf{u}^\top\frac{\partial\mathbf{M}}{\partial C_\kappa}\mathbf{Dv}\right]_n^{\kappa=1}$$

$$= \left[\mathbf{u}^\top(\mathbf{P}\boldsymbol{\pi}\,\mathbf{e}^\top - \mathbf{P})\mathbf{e}_\kappa D_\kappa\,v_\kappa\right]_n^{\kappa=1}$$

$$= \mathbf{u}^\top(\boldsymbol{\pi}\mathbf{v}^\top - \mathbf{PD}_\mathbf{v})\mathbf{D}.$$

4. Since $\mathcal{N} := \{\boldsymbol{\xi}: \nabla_\mathbf{C}\,\rho(\mathbf{MD})\,\boldsymbol{\xi} = 0\}$ is defined by a single constrain, it is an $n - 1$ dimensional linear subspace. Verification is given that $\boldsymbol{\zeta} = \mathbf{D}_v^{-1}\mathbf{D}^{-1}\boldsymbol{\pi} \in \mathcal{N}$:

$$\nabla_\mathbf{C}\,\rho(\mathbf{MD})\,\boldsymbol{\zeta} = \mathbf{u}^\top(\boldsymbol{\pi}\mathbf{v}^\top - \mathbf{PD}_\mathbf{v})\mathbf{D}(\mathbf{D}_v^{-1}\mathbf{D}^{-1}\boldsymbol{\pi})$$

$$= \mathbf{u}^\top(\boldsymbol{\pi}\mathbf{v}^\top\mathbf{DD}_v^{-1}\mathbf{D}^{-1}\boldsymbol{\pi} - \mathbf{P}\boldsymbol{\pi}) = \mathbf{u}^\top(\boldsymbol{\pi}\mathbf{v}^\top\mathbf{D}_v^{-1}\boldsymbol{\pi} - \boldsymbol{\pi})$$

$$= \mathbf{u}^\top(\boldsymbol{\pi}\mathbf{e}^\top\boldsymbol{\pi} - \boldsymbol{\pi}) = \mathbf{u}^\top(\boldsymbol{\pi} - \boldsymbol{\pi}) = 0.$$

∎

8 Discussion

There are two sets of take-home messages from the results here: one, content, and the other, methodology. Some of the content can be summarized simply as, "the results of

McNamara and Dall (2011) generalize to n environments," but in that generalization, new relationships emerge that were not visible from the parameters with only 2 environments.

The content in general may be understood to be part of the larger body of literature on the *Reduction Principle* and the departures from it, which include domains reaching from the evolution of recombination to the evolution of cultural traditionalism. The environmental change model of McNamara and Dall produces some specific new results for the reduction phenomena.

We see first that the "build up of the genotype on good sites" can be defined precisely as the *fitness-abundance covariance* — the covariance between the environment-specific growth rate and the excess abundance above what would emerge without differential growth. The phase of the census — whether taken before or after dispersal — is critical to the properties of the fitness-abundance covariance.

Exploration of the stationary state fitness-abundance covariance and its dependence on census phase is made for general combination of stochastic $\mathbf{M}$ and growth rates $\mathbf{D}$ in Theorem 3, and Corollaries 4, 11, 12. The only constraints are that $\mathbf{M}$ be irreducible and growth rates D_i be positive. Thus, they could just as well apply to mutation/selection balances as to dispersal/growth balances. Results for specific classes of $\mathbf{M}$ are found in Theorems 5, 6, and 7, and Corollary 13.

Theorem 3 shows that the process of dispersal decreases the fitness-abundance covariance by the variance in growth rates — a version of Fisher's Fundamental Theorem. Corollary 4 shows that the derivatives of post-dispersal $\mathcal{FA}(\mathbf{MD})$ and of $\rho(\mathbf{MD})$ always have the same sign with respect to any differentiation of $\mathbf{M}$. Theorem 5 finds that the pre-dispersal fitness-abundance covariance, $\mathcal{FA}(\mathbf{v}(\mathbf{DM}))$, is *always* positive when $\mathbf{M}$ is the transition matrix of an ergodic reversible Markov chain with all nonnegative eigenvalues, and growth rates differ between environments. Reversibility is important here, because a counterexample with $\mathcal{FA}(\mathbf{v}(\mathbf{DM})) < 0$ is found for periodic chains where one environment has a very small growth rate (Theorem 7). It is reasonable to conjecture, given the small region of growth rates in which $\mathcal{FA}(\mathbf{v}(\mathbf{DM})) < 0$, that $\mathcal{FA}(\mathbf{v}(\mathbf{DM})) > 0$ for all reversible chains regardless of the signs of their eigenvalues. Cyclic $\mathbf{M}$, on the other hand, always produce a negative post-dispersal fitness-abundance covariance, $\mathcal{FA}(\mathbf{MD})$ (Theorem 6).

When the growth rate of an environment is increased, then its stationary proportion of the population increases when the census is just *prior* to dispersal (Corollary 11). When the population is censused just *after* dispersal, the relationship can be reversed, as McNamara and Dall (2011) discovered, by extreme patterns of environmental change. Thus, we have a novel implication, for populations near their stationary distribution, that comparison

of the abundance relationships before and after dispersal can provide information about the extremity of the environmental change pattern (Corollary 13).

A number of results are obtained for a generalization of the McNamara and Dall (2011) model to n environments (7). Corollary 14 finds , just as McNamara and Dall do for two environments, that the post-dispersal fitness-abundance covariance, $\mathcal{FA}(\mathbf{MD})$ (which McNamara and Dall call the "multiplier effect"), is positive exactly when the reduction principle operates — i.e. when the growth rate of the population increases from reduced unconditional dispersal. It is negative when there are departures from reduction. This correspondence between a negative fitness-abundance covariance and departures from the reduction phenomenon is, however, specific to the model of McNamara and Dall and not a general property of departures from reduction for operators of the form $\mathbf{MD}$.

The field ecologist would want to know how feasible it is to measure the fitness-abundance covariance. Recall that $\mathbf{M}$ can represent a variety of processes. When $\mathbf{M}$ represents the dispersal probabilities between patches, then v_i represents the portion of the population in patch i, and the quantity π_i represents the portion that patch i *would have* in the absence of differential growth rates. Thus π_i is not something that actually exists but is a counterfactual. It may be feasible, however, to estimate $\boldsymbol{\pi}$ by estimating $\mathbf{M}$ from a measurement of the amount of dispersal between each patch (e.g. through mark and recapture experiments), and computing the Perron vector of the resulting estimated $\mathbf{M}$.

$\mathbf{M}$ has a different meaning in the model of McNamara and Dall, where it represent the Markov chain that the environmental states independently follow in all the patches, and π_i is simply the portion of patches that are in environmental state i, while v_i is the portion of the population in patches of environmental state i. Each of these is an actual quantity that is potentially measurable.

The expression $\tau_1^{-1} + \tau_2^{-1}$ from McNamara and Dall (2011) is seen in the general case to be a part of the harmonic mean of the expected durations of states in a Markov chain. The harmonic mean is shown in Lemma 15 to be a simple function of the sum of the eigenvalues of the chain's transition matrix. Thus the condition on $\tau_1^{-1} + \tau_2^{-1}$ discovered by McNamara and Dall is really a condition on the eigenvalues of the environment transition matrix.

In Theorem 16, these three entities — the reduction phenomenon, the harmonic mean of environment durations, and the eigenvalues of the environment transition matrix — are tied together in the case of environmental change processes that are reversible Markov chains. A sufficient condition for departures from reduction (selection for *increased* unconditional dispersal) is that all of the non-Perron eigenvalues of the environment transition matrix be negative, which represents an extreme pattern

of change, in which the harmonic mean of environment durations is less than $1 + 1/(n-1)$, where n is the number of environments. This means the environment changes almost every generation.

This departure from reduction identified by McNamara and Dall and generalized in Theorem 16 provides a new example summarized by the "principle of partial control" (Altenberg, 1984). The 'partial control' in Theorem 16 is that while the organism can control the transformation of its location (i.e. dispersal), it cannot control the transformations that change its environment.

Theorem 16 shows that a sufficient condition for the reduction phenomenon (selection for reduced unconditional dispersal) is that all eigenvalues of the environment transition matrix be positive, corresponding to less extreme environmental change. A general treatment of the intermediate case — of mixed positive and negative eigenvalues — remains an open question. But Corollary 19 shows that there is always some intermediate level of environmental change below which the reduction principle operates. The reduction principle would be expected to operate for more common patterns of environmental change.

Theorem 20 shows that this complexity of behavior disappears when the process of environmental change does not have any causal connection between the identity of sequential environments. In this case, only the reduction principle operates.

McNamara and Dall's model of conditional dispersal is here generalized to arbitrary numbers of environments. Theorem 21 shows that conditional dispersal provides another situation where we observe departures from the reduction principle. Conditional dispersal is mathematically analogous to directed mutation. Theorem 21 finds that there is always some environment from which it pays to increase dispersal, provided that there is: 1) some level of environmental change, 2) a causal connection between the current and next environments, and 3) different growth rates among environments. Therefore, philopatry is not the global evolutionarily stable state. This result holds for arbitrary environmental change Markov chains, and holds whether or not *unconditional* dispersal follows the reduction principle.

This seems to contradict the conclusion of McNamara and Dall (2011) that there are "conditions under which reliable, cost-free cues to habitat quality, which might intuitively influence optimal dispersal decisions, should be ignored in favour of blind natal philopatry." This contradiction is resolved by examining the complete adaptive landscape for the conditional dispersal rates (the $n = 2$ case in Figure 1).

We see that the evolutionarily stable state of dispersal is highly sensitive to any genetic or phenotypic constraints placed on the range of dispersal combinations. The hypothesized error rate for environmental cues in the McNamara and Dall model can constrain the variation to a region where the population growth rate is maximized by philopatry. But a slight decrease in the error rate can shift the evolutionarily stable state to maximize conditional dispersal from one environment, as shown by McNamara and Dall.

Other patterns of phenotypic constraint can be envisioned, and the sensitivity of the ESS in this model to phenotypic constraints leads to a diversity of potential phenomena: intermediate ESS states, bimodal states, or a general condition of evolutionary volatility. The evolutionary outcome becomes highly dependent on the variational properties (Altenberg, 1995) of the organism. To the extent that the McNamara and Dall model of random environments applies to the real world, the results suggest that empirical studies of the evolution of dispersal should find volatile relationships between an organism's dispersal behavior, the variational properties of its dispersal phenotype, and the pattern of environmental change its lineage has experienced.

8.1 Mathematical Methods

The second set of take-home messages from this paper regards the mathematical methods. The primary message is that techniques from the Reduction Principle literature and contemporary linear algebra allow one to obtain analytical results in greater generality than is often pursued. The common restriction to 2×2 matrices can be dropped for many results.

There is the added benefit from generalizing 2×2 models to the $n \times n$ case, which is that one is forced to see beyond the four particular entries of the 2×2 matrices to their deeper underlying structures, in particular their eigenvalues and eigenvectors, covariances, and the variational structure of the matrices. In the case of McNamara and Dall (2011, online Appendix A, Theorem A), the set of inequalities on the particular vector elements can be unified by a single inequality on a covariance expression, as in Corollary 14. It is hoped that the tractability of many results for general n, and the insights provided from such results, will encourage this approach more widely.

Tractability for Theorems 5, 16, 20, and 23, and Corollaries 17, 18, and 19 requires the assumption that the environments form a reversible Markov chain. The transition matrices of reversible Markov chains are synonymous with symmetrizable stochastic matrices. The tractability provided by symmetrizable stochastic matrices is the key tool adopted from Karlin's Theorem 5.1 (1982) and Friedland and Karlin (1975, Theorem 4.1). Karlin's Theorem 5.1 appears to have never been used since its publication until it was applied to the analysis of the evolution of mutation rates at multiple loci in (Altenberg, 2009a). Recently, however, symmetrizable stochastic matrices have been used by Schreiber and Li (2011) to analyze the evolution of dispersal in cyclic environments.

The environmental cycling that produces a departure

from the reduction principle in the model of Schreiber and Li (2011) satisfies the same condition of extreme environmental change as in Theorem 16, and the matrices are symmetrizable as well. But it is fundamentally a different model in that the environments change synchronously throughout all the patches, not independently as in (7), so it is represented by (9), $\mathbf{z}(t+2) = \mathbf{MD}_{(2)}\mathbf{MD}_{(1)}\mathbf{z}(0)$. Nevertheless, the parallels in its behavior with that of the McNamara and Dall model are intriguing. Recall that Karlin (1982) represented periodic environments by using cyclic matrices, so the model of Schreiber and Li (2011) can be represented as

$$\begin{bmatrix} \mathbf{0} & \mathbf{M}(m) \\ \mathbf{M}(m) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{D}_{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{(2)} \end{bmatrix}$$
$$= \left((1-m) \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} + m \begin{bmatrix} \mathbf{0} & \mathbf{P} \\ \mathbf{P} & \mathbf{0} \end{bmatrix} \right) \begin{bmatrix} \mathbf{D}_{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{(2)} \end{bmatrix}.$$

We note that this has the form $(1-m)\mathbf{A} + m\mathbf{B}$ (4) from the open problem posed in Altenberg (2004), and therefore provides another set of conditions on $\mathbf{A}$ and $\mathbf{B}$ that produce departures from reduction.

The case of general Markov chains remains an open problem for the above results. The principle difference when considering general Markov chains is that the non-Perron eigenvalues may come in complex-conjugate pairs, which represent cycles of states that are more probable in one direction than the reverse. Whether directional cycles of the environments can produce any new phenomena for the evolution of dispersal is here an open question.

8.2 Conclusions

Andrewartha (1961) classically defined ecology as "the scientific study of the distribution and abundance of organisms." In this respect, the fitness-abundance covariance investigated here is a basic quantity for ecology.

What makes its behavior more complex than intuition would suggest is that differential growth rates between patches or environments can interact with the multitude of possible dispersal, environmental change, and other mixing processes to produce novel relationships. The relationships identified by McNamara and Dall (2011) between the fitness-abundance covariance, the temporal properties of environmental change, and selection for or against dispersal provided the motivation for the present study.

The goal here has been to pursue the mathematics underlying these relationships. In so doing, these relationships are shown to connect to the body of work in the population genetics literature on the Reduction Principle for the evolution of genetic systems and migration, and provide new examples of departure from reduction. The common mathematics underlying all of these models may lead to the eventual development of a unified theoretical treatment in which the different ecological and evolutionary phenomena are seen as different aspects of a single phenomenological structure.

References

Ababneh, F., Jermiin, L. S., and Robinson, J. 2006. Generation of the exact distribution and simulation of matched nucleotide sequences on a phylogenetic tree. *Journal of Mathematical Modelling and Algorithms*, 5(3):291–308.

Altenberg, L. 1984. *A Generalization of Theory on the Evolution of Modifier Genes.* Ph.D. dissertation, Stanford University. http://dynamics.org/Altenberg/FILES/Altenberg_PhD.1984.pdf.

Altenberg, L. 1995. Genome growth and the evolution of the genotype-phenotype map. In Banzhaf, W. and Eeckman, F. H., editors, *Evolution and Biocomputation: Computational Models of Evolution*, volume 899 of *Lecture Notes in Computer Science*, pages 205–259. Springer-Verlag, Berlin.

Altenberg, L. 2004. Open problems in the spectral analysis of evolutionary dynamics. In Menon, A., editor, *Frontiers of Evolutionary Computation*, volume 11 of *Genetic Algorithms And Evolutionary Computation Series*, pages 73–102. Kluwer Academic Publishers, Boston, MA.

Altenberg, L. 2009a. Evolutionary reduction of mutation rates in a multivariate, multilocus model. *arXiv.org*, http://www.arxiv.com/abs/0909.2454v1.

Altenberg, L. 2009b. The evolutionary reduction principle for linear variation in genetic transmission. *Bulletin of Mathematical Biology*, 71(5):1264–1284.

Altenberg, L. 2010. Karlin theory on growth and mixing extended to linear differential equations. *arXiv.org*, http://arxiv.org/abs/1006.3147.

Altenberg, L. 2011. An evolutionary reduction principle for mutation rates at multiple loci. *Bulletin of Mathematical Biology*, 73:1227–1270.

Altenberg, L. and Feldman, M. W. 1987. Selection, generalized transmission, and the evolution of modifier genes. I. The reduction principle. *Genetics*, 117:559–572.

Andrewartha, H. G. 1961. *Introduction to the Study of Animal Populations.* University of Chicago Press.

Asmussen, M. 1983. Evolution of dispersal in density regulated populations: a haploid model. *Theoretical Population Biology*, 23(3):281–299.

Balkau, B. and Feldman, M. W. 1973. Selection for migration modification. *Genetics*, 74:171–174.

Barton, N. and Otto, S. 2005. Evolution of recombination due to random drift. *Genetics*, 169(4):2353.

Bergman, A. and Feldman, M. 1990. More on selection for and against recombination. *Theoretical Population Biology*, 38(1):68–92.

Blanquart, F. and Gandon, S. 2010. Evolution of migration in a periodically changing environment. *The American Naturalist*, 177(2):188–201.

Cairns, J., Overbaugh, J., and Miller, S. 1988. The origin of mutants. *Nature*, 335(6186):142–145.

Caswell, H. 2000. *Matrix Population Models*. Sinauer Associates, 2nd edition.

Cavalli-Sforza, L. and Feldman, M. W. 1973. Models for cultural inheritance. I. Group mean and within group variation. *Theoretical Population Biology*, 4:42–55.

Charlesworth, B. 1976. Recombination modification in a fluctuating environment. *Genetics*, 83:181–195.

Charlesworth, B. 1990. Mutation–selection balance and the evolutionary advantage of sex and recombination. *Genetical Research (Cambridge)*, 55:199–221.

Charlesworth, B. and Charlesworth, D. 1979. Selection on recombination in clines. *Genetics*, 91:581–589.

Charlesworth, B., Charlesworth, D., and Strobeck, C. 1979. Selection for recombination in partially self-fertilizing populations. *Genetics*, 93:237–244.

Cowen, R., Paris, C., and Srinivasan, A. 2006. Scaling of connectivity in marine populations. *Science*, 311(5760):522.

Donsker, M. D. and Varadhan, S. R. S. 1975. On a variational formula for the principal eigenvalue for operators with maximum principle. *Proceedings of the National Academy of Sciences U.S.A.*, 72:780–783.

Eigen, M. and Schuster, P. 1977. The hypercycle: A principle of natural self-organization. *Naturwissenschaften*, 64:541–565. 10.1007/BF00450633.

Elsner, L., Johnson, C., and Neumann, M. 1982. On the effect of the perturbation of a nonnegative matrix on its Perron eigenvector. *Czechoslovak Mathematical Journal*, 32:99–109.

Eshel, I. 1973a. Clone–selection and optimal rates of mutation. *Journal of Applied Probability*, 10:728–738.

Eshel, I. 1973b. Clone selection and the evolution of modifying features. *Theoretical Population Biology*, 4:196–208.

Feldman, M. W. 1972. Selection for linkage modification: I. Random mating populations. *Theoretical Population Biology*, 3:324–346.

Feldman, M. W. 1980. Personal communication.

Feldman, M. W. and Balkau, B. 1972. Some results in the theory of three gene loci. In Greville, T. N. E., editor, *Population Dynamics*, pages 357–383. Academic Press, New York.

Feldman, M. W. and Balkau, B. 1973. Selection for linkage modification II. A recombination balance for neutral modifiers. *Genetics*, 74:713–726.

Feldman, M. W., Christiansen, F. B., and Brooks, L. D. 1980. Evolution of recombination in a constant environment. *Proceedings of the National Academy of Sciences U.S.A.*, 77:4838–4841.

Feldman, M. W. and Krakauer, J. 1976. Genetic modification and modifier polymorphisms. In Karlin, S. and Nevo, E., editors, *Population Genetics and Ecology*, pages 547–583. Academic Press, New York.

Feldman, M. W. and Liberman, U. 1986. An evolutionary reduction principle for genetic modifiers. *Proceedings of the National Academy of Sciences U.S.A.*, 83:4824–4827.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fisher, R. A. 1958. *The Genetical Theory of Natural Selection*. Dover, New York.

Fretwell, S. 1972. *Populations in a seasonal environment*, volume 5 of *Monographs in Population Biology*. Princeton University Press.

Fretwell, S. and Lucas, H. 1969. On territorial behavior and other factors influencing habitat distribution in birds. *Acta biotheoretica*, 19(1):16–36.

Friedland, S. and Karlin, S. 1975. Some inequalities for the spectral radius of non-negative matrices and applications. *Duke Mathematical Journal*, 42(3):459–490.

Geweke, J., Marshall, R., and Zarkin, G. 1986. Mobility indices in continuous time markov chains. *Econometrica: Journal of the Econometric Society*, pages 1407–1423.

Gillespie, J. 1981a. Evolution of the mutation rate at a heterotic locus. *Proceedings of the National Academy of Sciences U.S.A.*, 78(4):2452.

Gillespie, J. 1981b. The role of migration in the genetic structure of populations in temporally and spatially varying environments. iii. migration modification. *The American Naturalist*, 117(3):223–233.

Haldane, J. 1937. The effect of variation of fitness. *The American Naturalist*, 71(735):337–349.

Hall, B. 1990. Spontaneous point mutations that occur more often when advantageous than when neutral. *Genetics*, 126(1):5–16.

Hastings, A. 1983. Can spatial variation alone lead to selection for dispersal? *Theoretical Population Biology*, 24(3):244–251.

Holsinger, K., Feldman, M. W., and Altenberg, L. 1986. Selection for increased mutation rates with fertility differences between matings. *Genetics*, 112:909–922.

Holsinger, K. E. and Feldman, M. W. 1983a. Linkage modification with mixed random mating and selfing: a numerical study. *Genetics*, 103:323–333.

Holsinger, K. E. and Feldman, M. W. 1983b. Modifiers of mutation rate: Evolutionary optimum with complete selfing. *Proceedings of the National Academy of Sciences U.S.A.*, 80:6732–6734.

Horn, R. A. and Johnson, C. R. 1985. *Matrix Analysis*. Cambridge University Press, Cambridge.

Ishii, K., Matsuda, H., Iwasa, Y., and Sasaki, A. 1989. Evolutionarily stable mutation rate in a periodically changing environment. *Genetics*, 121:163–174.

Karlin, S. 1976. Population subdivision and selection migration interaction. In Karlin, S. and Nevo, E., editors, *Population Genetics and Ecology*, pages 616–657. Academic Press, New York.

Karlin, S. 1982. Classifications of selection–migration structures and conditions for a protected polymorphism. In Hecht, M. K., Wallace, B., and Prance, G. T., editors, *Evolutionary Biology*, volume 14, pages 61–204. Plenum Publishing Corporation.

Karlin, S. and Carmelli, D. 1975. Numerical studies on two-loci selection models with general viabilities. *Theoretical Population Biology*, 7(3):399.

Karlin, S. and McGregor, J. 1972. The evolutionary development of modifier genes. *Proceedings of the National Academy of Sciences U.S.A.*, 69:3611–3614.

Karlin, S. and McGregor, J. 1974. Towards a theory of the evolution of modifier genes. *Theoretical Population Biology*, 5:59–103.

Keightley, P. and Otto, S. 2006. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature*, 443(7107):89–92.

Keilson, J. 1979. *Markov Chain Models: Rarity and Exponentiality*. Springer-Verlag, New York.

Kessler, D. A. and Levine, H. 1998. Mutator dynamics on a smooth evolutionary landscape. *Physical Review Letters*, 80(9):2012–2015.

Kimura, M. 1956. A model of a genetic system which leads to closer linkage by natural selection. *Evolution*, 10:278–287.

Kimura, M. 1967. On the evolutionary adjustment of spontaneous mutation rates. *Genetical Research (Cambridge)*, 9:23–34.

Kingman, J. F. C. 1961. A mathematical problem in population genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 57:574–582.

Kingman, J. F. C. 1978. A simple model for the balance between selection and mutation. *Journal of Applied Probability*, 15:1–12.

Kingman, J. F. C. 1980. *Mathematics of Genetic Diversity*. Society for Industrial and Applied Mathematics, Philadelphia.

Kirkland, S., Li, C. K., and Schreiber, S. J. 2006. On the evolution of dispersal in patchy landscapes. *SIAM Journal on Applied Mathematics*, 66:1366–1382.

Kondrashov, A. S. 1982. Selection against harmful mutations in large sexual and asexual populations. *Genetical Research (Cambridge)*, 40:325–332.

Kondrashov, A. S. 1984. Deleterious mutations as an evolutionary factor. I. The advantage of recombination. *Genetical Research (Cambridge)*, 44:199–217.

Kouyos, R., Silander, O., and Bonhoeffer, S. 2007. Epistasis between deleterious mutations and the evolution of recombination. *Trends in Ecology & Evolution*, 22(6):308–315.

Lenski, R. E. and Mittler, J. E. 1993. The directed mutation controversy and neo–Darwinism. *Science*, 259(5092):188–194.

Levinton, J. 1988. *Genetics, Paleontology, and Macroevolution*. Cambridge University Press, Cambridge. pp. 224–225, 494.

Liberman, U. and Feldman, M. W. 1986a. A general reduction principle for genetic modifiers of recombination. *Theoretical Population Biology*, 30:341–371.

Liberman, U. and Feldman, M. W. 1986b. Modifiers of mutation rate: A general reduction principle. *Theoretical Population Biology*, 30:125–142.

Liberman, U. and Feldman, M. W. 1989. The reduction principle for genetic modifiers of the migration rate. In Feldman, M. W., editor, *Mathematical Evolutionary Theory*, pages 111–137. Princeton University Press.

McNamara, J. M. and Dall, S. R. 2011. The evolution of unconditional strategies via the 'multiplier effect'. *Ecology Letters*, 14(3):237–243.

Muller, H. 1950. Our load of mutations. *American Journal of Human Genetics*, 2(2):111.

Nei, M. 1967. Modification of linkage intensity by natural selection. *Genetics*, 57:625–641.

Nei, M. 1969. Linkage modification and sex difference in recombination. *Genetics*, 63(3):681.

Odling-Smee, J. 2007. Niche inheritance: A possible basis for classifying multiple inheritance systems in evolution. *Biological Theory*, 2(3):276–289.

Olivieri, I., Michalakis, Y., and Gouyon, P. 1995. Metapopulation genetics and the evolution of dispersal. *The American Naturalist*, 146(2):202–228.

Otto, S. P. and Feldman, M. W. 1997. Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theoretical Population Biology*, 51:134–147.

Prais, S. 1955. Measuring social mobility. *Journal of the Royal Statistical Society. Series A (General)*, 118(1):56–66.

Price, G. R. 1972. Extension of covariance selection mathematics. *Annals of Human Genetics*, 35:485–489.

Pulliam, H. 1988. Sources, sinks, and population regulation. *The American Naturalist*, 132(5):652–661.

Pylkov, K. V., Zhivotovsky, L. A., and Feldman, M. W. 1998. Migration versus mutation in the evolution of recombination under multilocus selection. *Genetical Research (Cambridge)*, 71:247–256.

Ross, S. M. 1983. *Stochastic Processes*. John Wiley and Sons, New York.

Roze, D. and Barton, N. 2006. The hill–robertson effect and the evolution of recombination. *Genetics*, 173(3):1793.

Sasaki, A. and Iwasa, Y. 1987. Optimal recombination rate in fluctuating environments. *Genetics*, 115:377–388.

Schauber, E., Goodwin, B., Jones, C., and Ostfeld, R. 2007. Spatial selection and inheritance: applying evolutionary concepts to population dynamics in heterogeneous space. *Ecology*, 88(5):1112–1118.

Schreiber, S. and Li, C. 2011. Evolution of unconditional dispersal in periodic environments. *Journal of Biological Dynamics*, 5(2):120–134.

Schreiber, S. J. and Lloyd-Smith, J. O. 2009. Invasion dynamics in spatially heterogeneous environments. *The American Naturalist*, 174(4):490–505.

Seneta, E. 2006. *Non-negative Matrices and Markov Chains*. Springer-Verlag, New York, revised printing edition.

Shorrocks, A. F. 1978. The measurement of mobility. *Econometrica*, 46(5):1013–1024.

Steele, J. M. 2004. *The Cauchy-Schwarz Master Class*. Cambridge University Press.

Steinmeyer, S. H. and Wilke, C. O. 2009. Lethal mutagenesis in a structured environment. *Journal of Theoretical Biology*, 261:67–73.

Teague, R. 1977. A model of migration modification. *Theoretical Population Biology*, 12:86–94.

Tuljapurkar, S. 1990. *Population dynamics in variable environments*. Springer-Verlag New York.

Twomey, M. and Feldman, M. 1990. Mutation modification with multiplicative fertility selection. *Theoretical Population Biology*, 37(2):320–342.

University of Exeter 2011. 'Stupid strategies' could be best for the genes. *ScienceDaily*, March 1. `http://www.sciencedaily.com/releases/2011/02/110228104113.htm`.

Uyenoyama, M. and Waller, D. 1991a. Coevolution of self-fertilization and inbreeding depression I. mutation-selection balance at one and two loci. *Theoretical Population Biology*, 40(1):14–46.

Uyenoyama, M. and Waller, D. 1991b. Coevolution of self-fertilization and inbreeding depression II. symmetric overdominance in viability. *Theoretical Population Biology*, 40(1):47–77.

Wagner, G. P. and Altenberg, L. 1996. Complex adaptations and the evolution of evolvability. *Evolution*, 50(3):967–976.

Wiener, P. and Feldman, M. W. 1991. The evolution of dispersal in a model of mixed selfing and random mating. *Evolution*, 45(7):1717–1726.

Wiener, P. and Feldman, M. W. 1993. The effects of the mating system on the evolution of migration in a spatially heterogeneous population. *Evolutionary Ecology*, 7(3):251–269.

Wiener, P. and Tuljapurkar, S. 1994. Migration in variable environments: exploring life-history evolution using structured population models. *Journal of Theoretical Biology*, 166(1):75–90.