Competitive exclusion and limiting similarity: A unified theory

Géza Meszénaa, Mats Gyllenbergb, Liz Pásztorc, Johan A.J. Metzd,e

aDepartment of Biological Physics, Eötvös University, Pázmány Péter sétány 1A, H-1117 Budapest, Hungary
bDepartment of Mathematics and Statistics, Rolf Nevanlinna Institute, University of Helsinki, FIN-00014 Helsinki, Finland
cDepartment of Genetics, Eötvös University, Pázmány Péter sétány 1C, H-1117 Budapest, Hungary
dInstitute of Biology, Leiden University, van der Klaauw Laboratory, P.O. Box 9516, NL-2300 RA Leiden, The Netherlands

Received 4 November 2004
Available online 21 October 2005

Abstract

Robustness of coexistence against changes of parameters is investigated in a model-independent manner by analyzing the feedback loop of population regulation. We define coexistence as a fixed point of the community dynamics with no population having zero size. It is demonstrated that the parameter range allowing coexistence shrinks and disappears when the Jacobian of the dynamics decreases to zero. A general notion of regulating factors/variables is introduced. For each population, its impact and sensitivity niches are defined as the differential impact on, and the differential sensitivity towards, the regulating variables, respectively. Either the similarity of the impact niches or the similarity of the sensitivity niches results in a small Jacobian and in a reduced likelihood of coexistence. For the case of a resource continuum, this result reduces to the usual “limited niche overlap” picture for both kinds of niche. As an extension of these ideas to the coexistence of infinitely many species, we demonstrate that Roughgarden’s example for coexistence of a continuum of populations is structurally unstable.

1. Introduction

We expect the struggle for life between the kinds attempting to occupy the same niche and peaceful coexistence between the species established in different niches (Gause, 1934). This principle of “competitive exclusion” has become a cornerstone of ecological thinking while at the same time it has remained highly controversial. The basic models were introduced by MacArthur and Levins in the mid-1960s both for discrete and for continuous resources. These models differ considerably in their mathematical setup and sophistication.

The discrete models (MacArthur and Levins, 1964; Tilman, 1982) consider competition for a finite number of distinct resources and state that the number of coexisting species should not exceed the number of resources they compete for. This conclusion has a sound mathematical foundation: to have a structurally stable solution, i.e., a solution that does not disappear on the slightest change of the model specification, the number of equations describing the population dynamical equilibrium should not exceed the number of unknowns.

With some risk of becoming tautological, we can relax the assumption of resource competition by counting all the factors behaving like resources (Levin, 1970; Armstrong and McGehee, 1980; Heino et al., 1997). Limited practical usefulness is the price for theoretical robustness. It is a rare biological situation where the resources, or the regulating factors, are easy to count and low in number. In most cases, very many environmental factors that are potentially regulating are present. Which are the really important ones? Which of them should be considered as distinct from the others (cf. Abrams, 1988)? On the other hand, if only
the limiting resources are counted, their number often turns out to be too low to explain species diversity in a constant environment (Hutchinson, 1959).

The classical continuous model (MacArthur and Levins, 1967) studies the partitioning of a continuous scale of resources, e.g., seeds of different sizes. In this case, strictly speaking, an infinite number of different resources are present, i.e., each seed size has to be considered as a different resource. Consequently, the argument used for the discrete situation cannot be applied to bound the number of species in the continuous case. Still, we do not expect an infinite number of species to coexist. The classical concept of “limiting similarity” (Hutchinson, 1959), based on the study of the Lotka–Volterra competition model (MacArthur and Levins, 1967), states that the resource scale is partitioned between the species. The width of the “resource utilization function” of a species is expected to set the width of a single partition, referred to as the “niche breadth.” The allowed similarity of the coexisting species is limited and their number is bounded by the number of possible partitions. It seems to be reasonable to consider one partition of the resource scale as a single resource, distinct from the rest. Unfortunately, no mathematical theory has appeared as yet that fully captures this intuition in a general way: again, the question is, to what extent should the resources differ to be counted as different? Neither has any general conclusion emerged when the later studies went beyond the original Lotka–Volterra framework. These studies actually resulted in a fading away of the hope of finding a model-independent lower bound to similarity (Abrams, 1983).

The status of the limiting similarity principle is unclear even for the original Lotka–Volterra model. May (1973, p. 158) found that the limit of similarity of two coexisting species can be arbitrarily small if their respective carrying capacities are chosen to be sufficiently similar. Yodzis (1989, p. 125) states that, contrary to the two species case, there is a strict lower bound to similarity for three species. Probably, the most drastic blow against limiting similarity occurred when Roughgarden (1979, pp. 534–536) provided an example of coexistence of a continuum of types in the Lotka–Volterra model. While the example was intended to describe the phenotype distribution within a single species, it can be interpreted in the context of species coexistence. An infinite number of different resources does allow coexistence of an infinite number of species. The example thus seems to violate the whole idea of limiting similarity (Maynard Smith and Szathmáry, 1995).

Even though limiting similarity and resource partitioning failed to earn the status of a mathematical theory, they have remained widely accepted concepts in ecology (Begon et al., 1996, p. 300). However, if limiting similarity were just an artifact of some specific mathematical models, we would not be allowed to use it as a basis of biological reasoning. Without limiting similarity, the practical relevance of competitive exclusion would be constrained to the simplest cases of population regulation (Rosenzweig, 1995, p. 127). If we could not safely assume competitive exclusion between the variants of the same species, even the Darwinian concept of natural selection would lose its basis. The goal of the present paper is to carry out the mathematical step from the solid ground of competitive exclusion in the discrete case to establish the general existence of limiting similarity in a well-defined sense.

The key issue of species coexistence is the necessity for mechanisms stabilizing it (Chesson, 2000b). This vantage point allows us to investigate the problem independent of specific model assumptions. We start from May’s observation (May, 1973, 1974) that the more similar the two species are, the more narrow the range of \( K_1/K_2 \) permitting coexistence (May, 1973; \( K_1 \) and \( K_2 \) stand for the two carrying capacities. See also a similar analysis by Vandermeer, 1975, which uses the intrinsic growth rates, instead of the carrying capacities as a reference.) We generalize this statement beyond its original framework of a Lotka–Volterra-type model and to an arbitrary number of species. Limiting similarity is interpreted as a shrinking likelihood of coexistence with increasing similarity. Especially, we demonstrate that the Roughgarden type of continuous coexistence, as it is called, generally becomes impossible on the slightest change of the model. To relate limiting similarity to resource usage, or to the regulating factors, and to the concept of niche, we will apply Leibold’s (1995) distinction between the two legs of the population–environment interaction. We restrict our mathematical analysis to fixed-point attractors in a constant environment.

The theory is presented in several steps. After introducing our central notions in Section 2, we demonstrate the basic issues of limiting similarity and establish the connection between the discrete and the continuous cases via a linear model in Section 3. Then, in Section 4, the nonlinear, model-independent generalization is presented. Finally, we discuss the wider context of our work, in Section 5. Background mathematics are summarized for the convenience of the interested reader in the appendix.

2. Basic concepts

2.1. The notion of robust coexistence

Populations coexist when the long-term growth rates of all of them are simultaneously zero, i.e., the births just compensate for the deaths in each of them. Coexistence of any given set of populations can be achieved in many parameter-rich models by adjusting the parameters. The problem of coexistence becomes non-trivial only when one rejects parameter fine-tuning and requires the population dynamical equilibrium point to exist for a considerable range of the parameters. In a given environment, the wider this range, the more likely the coexistence. We will refer to this property as robustness of coexistence. Robustness of coexistence is measured by the volume of the set of parameter values permitting coexistence. In special, but
Population regulation is mediated by the regulating variables collected in vector \( E \). Elements of \( E \) may affect the populations but are not affected by the populations, so they are not involved in the feedback loop. Turchin (2003, p. 398) uses the terms endogeneous and exogeneous factors for the elements of \( I \) and \( E \), respectively. The distinction between resources and conditions (Begon et al., 1996) is analogous, but see the Discussion.

As a matter of definition, we suppose that the vectors \( E \) and \( I \) together represent a complete description of the environment affecting the demographic parameters of the individuals. Specifically, all the interactions between the individuals are mediated through \( I \). Elements of the vectors \( I \) and \( E \) together are often referred to as limiting factors.

In the mathematical treatment we restrict our attention to unstructured populations and describe the state of the community by the population sizes \( n_1, n_2, \ldots, n_L \) of the \( L \) coexisting species. (See Section 5.4 for the discussion of the spatial structure of the community.) These sizes determine the values of the regulating variables:

\[
I : (n_1, n_2, \ldots, n_L) \mapsto I = (n_1, n_2, \ldots, n_L).
\]

The map \( I \) will be referred to as the impact function (or output map in the terminology of Diekmann et al. (2001, 2003)). In turn, the regulating variables, at a given \( E \), determine the population growth rates \( r_i \):

\[
S : I \mapsto (r_1(E, I), r_2(E, I), \ldots, r_L(E, I)).
\]

We will refer to the dependence relation \( S \) as the sensitivity function.

The composition of the impact and the sensitivity maps determines the population regulation \( R \) which is the dependence of the growth rates on the population sizes at fixed \( E \):

\[
R = S \circ I : (n_1, n_2, \ldots, n_L) \mapsto (r_1(E, I), r_2(E, I), \ldots, r_L(E, I)).
\]

The population regulation map determines the population dynamics according to the following:

\[
\frac{1}{n_i} \frac{dn_i}{dt} = r_i(E, I(n_1, n_2, \ldots, n_L)) \quad (i = 1, 2, \ldots, L).
\]

A simple calculation leads to

\[
\frac{d}{dt} \left( \frac{n_i}{n_j} \right) = (r_i - r_j) \frac{n_i}{n_j} \quad (i, j = 1, 2, \ldots, L)
\]

showing that the ratio of the sizes of the \( i \)th and the \( j \)th populations is governed by the difference between their growth rates.

---

**Fig. 1.** The general scheme of an environmental feedback loop. The dynamics of the populations are determined by the growth rates. They, in turn, are determined by the population sizes via population regulation \( R \). Population regulation is mediated by the regulating variables collected in the vector \( E \). The map \( R \) is a composition of the population impact \( I \) on the regulating variables and the population sensitivity \( S \) towards these variables. The external environmental parameters, collected in vector \( E \), also affect the growth rates. However, these parameters are not involved in the feedback loop as they are independent of the population sizes. The elements of \( E \) and \( I \), together, are often referred to as limiting factors.

important, situations this volume shrinks to zero and robustness is lost completely. Then, coexistence can be achieved only for special choices of the parameters. We will refer to such coexistence as structurally unstable. It has probability zero of occurring in the real word.

As we will see, robust coexistence is a consequence of feedbacks (or stabilizing mechanisms, Chesson, 2000b), to be called population regulation, with the potential of adjusting the birth and death rates (Fig. 1). The growth rates are nullified at different values of the external environmental parameters by different combinations of the population sizes. When the population regulation is weak, changing the population sizes within the acceptable (i.e., positive) range can compensate only for small changes of the external parameters. Therefore, the robustness of a particular pattern of coexistence is weak, or lost, when the underlying population regulation is weak, or lost.

**2.2. Decomposition of the feedback loop**

While resource limitation is the archetypical mechanism of population regulation, many other types of interactions, like predation, infection, etc., can be at the root of the unavoidable regulation. For a general treatment, it is useful to introduce the notion of regulating variables (or regulating factors, Case, 2000, p. 146; Krebs, 2001, p. 288).

The regulating variables, collected in the vector \( I \), are the variables mediating the feedback. Each element of the vector \( I \) depends on the population sizes and affects some of the demographic parameters of the individuals in the populations. Regulating factors are also referred to as “environmental interaction variables”, or just “environmental variables” (Heino et al., 1997; Diekmann et al., 2001, 2003; Vukics et al., 2003).
In a theoretical experiment, and sometimes in a real one, one can fix the regulating variables. Doing so, the feedback loop opens up, population regulation ceases to operate and the individuals reproduce and die independently. In this case the equations of population dynamics (4) become linear with fixed values of the growth rates \( r_i \). Then, the populations grow exponentially and the ratio of population sizes changes monotonically according to the differences in the growth rates [Eq. (5)]. For instance, if resource limitation creates the only interaction between the individuals, continuous replenishment of the resources leads to unlimited exponential growth of the now independent populations.

2.3. The discrete bound of diversity

The set of equilibrium equations

\[ r_i(E,I) = 0 \quad (i = 1, 2, \ldots, L) \]  

(6)

consists of \( L \) equations for \( \dim I \) unknowns. Generically, they can be solved only if \( L \leq \dim I \). A solution may exist for \( L > \dim I \), but such a solution is structurally unstable. If the smallest set of regulating variables is chosen,

\[ D = \dim I \]  

(7)

can be regarded as the dimension of population regulation (Meszéna and Metz, 1999; Heino et al., 1997; Diekmann et al., 2003). Using this quantity, one can state the following.

Discrete Competitive Exclusion Principle. A structurally stable coexistence of \( L \) populations requires

\[ L \leq D \]  

(8)

where \( D \) is the dimension of population regulation.

A bound of this type is often referred to as the competitive exclusion principle. Without stressing the difference between the limiting and the regulating factors, its generality was recognized already by Levin (1970). (See also Armstrong and McGehee, 1980; Heino et al., 1997; Meszéna and Metz, 1999.) We use the qualifier discrete to distinguish from the more general meaning of the same term (cf. Hardin, 1960).

In the case of pure resource competition, i.e., when resource shortage is the only source of interaction between individuals, the quantities of the resources are the regulating variables and \( D \) is the number of different resources. As the elements of \( E \) are parameters, rather than unknowns, of the equilibrium equations, they do not count in this context. If a change in the environment does not influence the feedback loop it does not affect the maximal number of coexisting species, either.

As a corollary, no robust coexistence is possible for \( D = 1 \) (Metz et al., 1996b). If all populations are limited by the same resource, the species with the lowest equilibrium resource concentration on its own outcompetes all others (\( R^* \) rule, Tilman, 1982, p. 43). If all populations affect the resource equally, the sum of the population sizes can be regarded as the sole regulating variable (density-dependent, or \( K \) selection, MacArthur, 1962).

As the infinite number of different resources in a resource continuum demonstrates, the dimension of regulation is not necessarily finite. The discrete version of competitive exclusion is useless in this case.

3. Linear population regulation

3.1. Basics of the linear model

In this section we investigate the coexistence of \( L \) species in the case of linear regulation. As an arbitrary choice, let the origin \( I = 0 \) of the space of regulating variables correspond to the absence of the populations of our interest.

Assume that the population sizes affect \( I \) linearly

\[ (I) \quad I = n_1 C_1 + n_2 C_2 + \cdots + n_L C_L. \]  

(9)

Here, the \( D \)-dimensional vectors \( C_1, C_2, \ldots, C_L \) denote the per capita environmental impact of the corresponding populations. In case of pure resource competition, the components of \( I \) measure the depletion of the different resources caused by the presence of the populations.

Moreover, we also suppose that the growth rates depend linearly on the regulating variables, e.g., on the resource depletions

\[ (S) \quad r_i = r_{0i}(E) - S_i \cdot I \quad (i = 1, 2, \ldots, L), \]  

(10)

where the “\( \cdot \)” denotes a scalar product. \( r_{0i}(E) \) is the innate capacity of increase, which depends only on the external parameters \( E \). The \( D \)-dimensional vectors \( S_1, S_2, \ldots, S_L \) describe the sensitivity of the populations to the regulating variables. The minus sign was introduced to harmonize with the depletion interpretation of \( I \) for resource competition.

The two vectors related to the \( i \)th population, \( C_i \) and \( S_i \), characterize the interaction of this population with the regulating factors. These vectors will be referred to as impact and sensitivity niche vectors, respectively. They are related to, but not identical to, Leibold’s (1995) concept of impact and requirement niches. See the Discussion regarding connection to Hutchinson’s (1978) notion of niche.

For finite-dimensional regulation, the \( k \)th component \( C_{ik} \) of the impact niche vector \( C_i \) measures the per capita impact of the \( i \)th population on the \( k \)th regulating variable \( I_k \). Similarly, \( S_{ik} \) measures the sensitivity of this population towards \( I_k \). We will examine the infinite-dimensional case in Section 3.4.

Eqs. (9) and (10) together lead to a Lotka–Volterra-type population regulation

\[ (R) \quad r_i(E,I(n_1,n_2,\ldots,n_L)) = r_{0i}(E) - \sum_{j=1}^{L} a_{ij} n_j \]  

(11)

for each \( i \), where

\[ a_{ij} = S_i \cdot C_j \quad (i,j = 1, 2, \ldots, L) \]  

(12)
are the elements of the community matrix \( a \). Positive elements of \( a \) represent competitive interactions. The competition coefficients used in the conventional Lotka–Volterra formulation are \( a_{ij}/a_{ii} \) (see e.g., Vandermeer, 1975).

The set of equilibrium equations (6) can be solved uniquely, and in a structurally stable way, if the Jacobian determinant
\[
J = \det a
\]
of the system is different from zero. Then, the solution is
\[
n_i = \sum_{j=1}^{L} a_{ij}^{-1} r_{0j} = \frac{1}{J} \sum_{j=1}^{L} \text{adj}(a)_{ij} r_{0j} \quad (i = 1, 2, \ldots, L),
\]
where \( a_{ij}^{-1} \) and \( \text{adj}(a)_{ij} \) are the elements of the inverse and of the adjoint matrix of \( a \) (Anton, 1984, p. 80), respectively. (The adjoint matrix is not to be confused with the complex conjugate of the transposed matrix, which is usually also called the adjoint matrix.) The solution is biologically meaningful, i.e., corresponds to coexistence of populations only if all \( n_i \) are positive.

### 3.2. Strength of population regulation versus robustness of coexistence

Here we show that the robustness of coexistence (i.e., the volume of the permitting set of parameters) decreases zero gradually when \( J \to 0 \). Let
\[
U = \{(n_1, n_2, \ldots, n_L) | 0 < n_i < n_{\text{max}} \text{ for } i = 1, 2, \ldots, L \}
\]
be the set of admissible community states, i.e., the set of potential combinations of positive population sizes below the maximal \( n_{\text{max}} \), which can be arbitrarily high. Then,
\[
(r_{01}, r_{02}, \ldots, r_{0L}) \in aU
\]
(16)
specifies the combinations of \( r_{0j} \) values allowing coexistence, where \( aU \) represents the set of possible values of \( an \) when \( n \in U \). Using \( \mathcal{V} \) to denote volume,
\[
\mathcal{V}(aU) = |J|\mathcal{V}(U)
\]
(17)
(Apostol, 1962, pp. 84–86; Case, 2000, pp. 420–422, Edwards Jr., 1973, p. 245) which proves our claim. Fig. 2 illustrates the two-dimensional case, where \( aU \) is a parallelogram (\( n_{\text{max}} \) is not represented).

Eq. (17) is a simple consequence of the geometrical interpretation of a determinant: the linear operation defined by the matrix \( a \) maps the unit cube into a parallelepiped of volume \( |J| \). This volume serves as a measure of the strength of population regulation on the community level. It is large if the parallelepiped is wide in each direction, i.e., when each population size affects at least some growth rates differently enough and each growth rate is affected differently enough by at least some population sizes.

Obviously, whenever the population regulation is weak, the inverse dependence (related to the matrix \( a^{-1} \)) will be strong. If a large change of the population sizes leads to a small change of the growth rates, then the equilibrium population sizes will be very sensitive to the factors affecting the growth rates. This sensitivity is represented in Eq. (14) by the denominator \( J \). The extreme sensitivity of the population sizes, when \( |J| \) is small, implies that a small change of the parameters could drive some of the populations into extinction. That is, a weak population regulation leads to a non-robust coexistence.

In the limiting case of \( J = 0 \), robustness is entirely lost. The equilibrium equations are solvable only for special choices of \( r_{0j} \). Coexistence, if it exists, is structurally unstable for \( J = 0 \).

### 3.3. Limiting similarity of the niches

We wish to explore the loss of robustness, discussed in the previous paragraph, in terms of the niche vectors.
Obviously, population regulation becomes weak if either the environmental impact or the environmental sensitivity becomes weak.

The strength of the population regulation $\mathcal{R}$ was measured by the volume $|J|$ of the image of the unit cube under $\mathcal{R}$. We need an analogous measure for the impact map $I$ and for the sensitivity map $S$. The impact function $I$ maps the unit cube into the $L$-dimensional parallelepiped spanned by the impact niche vectors $C_i$, while the sensitivity function $S$ maps the unit cube into a parallelepiped spanned by the sensitivity niche vectors $S_i$. The volume of these parallelepipeds will be denoted by $V_C$ and $V_S$, respectively.

Note that the volumes $V_C$ and $V_S$ are not determinants, because the maps $I$ and $S$ operate between spaces of different dimensionality. The related mathematics are summarized in the appendix. In the main text, we rely on the clear intuitive notion of the volume in a self-contained way.

The following statement establishes the expected connection between the strength of impact and sensitivity on the one hand and the strength of the regulation on the other.

**Proposition: Regulation strength.**

$$|J| \leq V_S \cdot V_C,$$

i.e., either weak impact or weak sensitivity leads to diminished population regulation.

**Proof of Proposition.** Let $\hat{C}$ be a $D \times L$ matrix composed of the column vectors $C_i$ and $\hat{S}$ an $L \times D$ matrix composed of the row vectors $S_i$:

$$a = \hat{S} \cdot \hat{C}. \quad (19)$$

For $L = D$, they are square matrices and

$$|J| = |\det \hat{S}| \cdot |\det \hat{C}| = V_S \cdot V_C. \quad (20)$$

For $D > L$, observe that only the $D$-dimensional linear subspace, containing the $S_i$ vectors, is relevant. We use a prime to denote the projection onto this subspace. Obviously, $V_S = V_S$ and $V_C \leq V_C$. Then,

$$|J| = V_S \cdot V_C \leq V_S \cdot V_C.
$$

Finally, note that $|J| = V_C = V_S = 0$ for $D < L$.

The volumes $V_C$ and $V_S$ measure the dissimilarity of the impact and the sensitivity niche vectors, respectively, in the linear sense. If the vectors $C_i$ $(i = 1, 2, \ldots, L)$ are linearly dependent then $V_C = 0$. A small value of $V_C$ represents a situation in which the vectors are close to being linearly dependent. Specifically, the similarity of any two or more of the impact niche vectors leads to small $V_C$. Analogously, their similarity between the sensitivity niche vectors results in a small value of $V_S$ and their linear dependence results in $V_S = 0$. According to the Proposition, as robust coexistence requires a large enough value of $|J|$, it also requires large enough values for $V_S$ and $V_C$. It is not realistic to suppose that an extreme smallness of $V_C$, i.e., the similarity of the impact vectors, is compensated by an extremely large $V_S$, or vice versa.

Expressed differently, all the impact niche vectors and all the sensitivity niche vectors should be different enough; otherwise, population regulation is weak and coexistence is restricted to a narrow range of parameter values. In more qualitative terms, populations have to differ both in their effects on the regulating variables and in their dependence on these variables (cf. Abrams, 1988).

For $L = 2$, the volume $V_S$ reduces to the area of the parallelogram spanned by the sensitivity niche vectors:

$$V_S = |S_1| |S_2| \sin(S_1, S_2), \quad (22)$$

where $(S_1, S_2)$ denotes the angle between the vectors $S_1$ and $S_2$. It can be related to the scalar product

$$S_1 \cdot S_2 = \sum_{i=1}^{D} S_{1i} S_{2i} \quad (23)$$

of the niche vectors which is considered as the measure of “niche overlap” (Petraitis, 1989). The scalar product can be expressed (Anton, 1984, p. 184) as

$$S_1 \cdot S_2 = |S_1| |S_2| \cos(S_1, S_2), \quad (24)$$

so the volume and the niche overlap are related as

$$(V_S)^2 = |S_1|^2 |S_2|^2 - (S_1 \cdot S_2)^2. \quad (25)$$

If the normalizations $|S_1|$ and $|S_2|$ are kept constant, the volume $V_S$ is small when the overlap of the sensitivity niches is large. A similar statement applies to the $C$ vectors. Thus, a small overlap of both kinds of niche vectors is a requirement for robust coexistence.

This phenomenon is demonstrated for $L = D = 2$ in Fig. 4. The two-dimensional regulating variable is

$$I = n_A C_A + n_B C_B, \quad (26)$$

![Fig. 3. The volume measures of the population regulation.](image)
Fig. 4. (a) Equilibrium of two species, A and B, with two-dimensional linear population regulation. The axes represent the regulating variables; the origin is the ecological vacuum. The impact niche vectors \( C_A \) and \( C_B \) determine the change of the regulating variables as a consequence of adding a single individual to the corresponding population. Non-negative population sizes result in the values of the regulating variables within the wedge between these vectors (gray). The two solid lines represent the Zero Net Growth Isocline\-s (ZNGI, Tilman, 1982) corresponding to the two equilibrium equations, \( r_A = 0 \) and \( r_B = 0 \), of the two species. They are perpendicular to the sensitivity niche vectors \( S_A \) and \( S_B \). The intersection point of the ZNGIs (large dot) must lie within the gray region to describe a biologically realistic coexistence. (b) The equilibrium for several different values of \( r_{0B} \). Some of them lie outside the allowed region and correspond to negative population sizes. (c) When the impact niche vectors are very similar, the allowed region is very narrow and most of the equilibrium points lie outside of it. (d) Here the sensitivity niche vectors are the similar ones. While the allowed region remains wide, the coexistence of the two species is limited to a narrow range of \( r_{0B} \).

where indices A and B are used to distinguish the two populations. As the biological interpretation requires \( n_{A0}, n_{B0} \geq 0 \), the wedge between the impact niche vectors \( C_A \) and \( C_B \) represents the possible environmental conditions. The sensitivity niche vectors point in the direction of steepest decrease of the growth rate of the corresponding population in the space of regulating variables. The environmental conditions satisfying the equilibrium equation of any of the two populations lie on a straight line, the Zero Net Growth Isocline (ZNGI, Tilman, 1982), which is perpendicular to the sensitivity niche vector of the corresponding population. The crossing point of the two ZNGIs is the equilibrium point. It must lie within the wedge of allowable environmental conditions to represent a biologically relevant coexistence of the two populations. If the equilibrium point lies outside the wedge, it corresponds to a (formal) solution of the equilibrium equations in which one of the population sizes is negative. In the panes (b–d) a series of equilibrium points, corresponding to a series of different values of \( r_{0B} \), is plotted. Observe that only a very narrow range of the \( r_{0B} \) values allows coexistence if either the two impact niches [as in (c)], or the two sensitivity niches [as in (d)], are similar in direction.

3.4. Niche overlap in case of a resource continuum

The linear theory developed in the previous sections applies for infinite-dimensional regulation without essential modifications. As an example, we investigate the textbook case of a resource continuum.

Suppose that some (e.g., bird) populations compete for seeds of different sizes. Denote the seed size (more generally: the quality of the resource) by \( q \). As \( q \) is a continuous quantity, a change of notation is in order. While the depletion \( I_k \) of the \( k \)th resource was the \( k \)th interaction variable in the finite-dimensional case, depletion \( I(q) \) of the resource of quality \( q \) plays the same role for the case of resource continuum. The function \( q \mapsto I(q) \) substitutes for the vector \( I \). The different notation does not reflect any conceptual difference between the finite and the infinite-dimensional case as the finite-dimensional vector \( I = \{I_1, I_2, \ldots, I_D\} \) can also be seen as a function \( k \mapsto I_k \) defined on the set of \( \{1, 2, \ldots, D\} \).

Also, the niche vectors become functions in the case of a resource continuum. For the \( i \)th population, \( C_i(q) \) and \( S_i(q) \) measure the impact on, and the sensitivity towards, the resource of quality \( q \), respectively. The linear impact and sensitivity maps can be written as

\[
(\mathcal{I}) \quad I(q) = \sum_{j=1}^{L} C_j(q)n_j
\]

and

\[
(\mathcal{S}) \quad r_i = r_{0i} - \int S_i(q)I(q) \, dq \quad (i = 1, 2, \ldots, L).
\]

Then, again,

\[
(\mathcal{R}) \quad r_i(E, \mathcal{I}(n_1, n_2, \ldots, n_L)) = r_{0i}(E) - \sum_{j=1}^{L} a_{ij}n_j
\]

for each \( i \), where now the elements of the community matrix are given by

\[
a_{ij} = \int S_i(q)C_j(q) \, dq \quad (i, j = 1, 2, \ldots, L).
\]

For the continuous case the niche overlap is measured by the “overlap integral” of the niche functions, which are the infinite-dimensional analogues of the scalar product (23)

\[
S_i \cdot S_j = \int S_i(q)S_j(q) \, dq.
\]

\[
C_i \cdot C_j = \int C_i(q)C_j(q) \, dq.
\]

The theory of the previous sections applies: the higher the niche overlap, the smaller the parameter range allowing coexistence. We emphasize again that both the impact and
the sensitivity niche functions must differ, in the sense of a small overlap, for robust coexistence.

We may arrive at the intuitive notion of the resource utilization function (MacArthur and Levins, 1967), if we suppose that the impact and the sensitivity niche functions are proportional (cf. Petraitis, 1989)

\[ C_i(q) = aS_i(q). \]  

(33)

Then, the resource utilization function can be defined as

\[ u_i(q) = \sqrt{a} S_i(q) = \frac{1}{\sqrt{a}} C_i(q). \]  

(34)

With this identification, the community matrix (i.e., the competition coefficients for resource competition) assumes the usual form

\[ a_{ij} = \int u_i(q) u_j(q) \, dq, \]  

(35)

i.e., it is the overlap of the two resource utilization functions. The small overlap is the usual condition for robust coexistence (cf. May, 1973, 1974).

The notion of the resource utilization function expresses the intuition that a population affects, and depends on, the resources that it uses. However, there is no strong biological basis for the exact proportionality expressed by Eq. (33). It should be considered as no more than a convenient simplifying assumption that can be removed with little effort. Yodzis (1989, p. 119) provides a mechanistic underpinning of Eq. (35) by explicit modeling of the resource dynamics. However, the emerging resource utilization function is an artificial combination of factors, lacking clear biological content. It is easy to repeat this derivation in a more natural way in terms of the impact and the sensitivity functions.

3.5. Structural instability of Roughgarden’s example

The non-robustness of coexistence of similar populations hints that the coexistence of a continuum of populations cannot be robust, either. Following Gyllenberg and Meszéna (2005), here we show that Roughgarden’s (1979) example for continuous coexistence is structurally unstable.

The Lotka–Volterra-type population regulation [Eq. (29)] can be rewritten for the continuous density distribution as

\[ \mathcal{R} \quad r(y) = r_0(y) - \int a(y, x) n(x) \, dx. \]  

(36)

Here, \( r(y) \) and \( r_0(y) \) are the growth rate and the intrinsic growth rate of trait \( y \), respectively. The distribution of the populations along the trait axis \( x \) is described by the density function \( n(x) \). As Roughgarden noted, the Gaussian choices

\[ r_0(x) = \frac{1}{\sqrt{2\pi w}} e^{-x^2/2w^2}, \]  

(37)

and

\[ a(y, x) = \frac{1}{\sqrt{2\pi \sigma}} e^{- (y - x)^2/2\sigma^2}. \]  

(38)

allow the continuous distribution

\[ n(x) = \frac{1}{\sqrt{2\pi (w^2 - \sigma^2)}} e^{-x^2/(2(w^2 - \sigma^2))}. \]  

(39)

of populations to coexist in equilibrium (Fig. 5a).

However, a small periodic perturbation of the intrinsic growth rate

\[ r'_0(x) = r_0(x)(1 + \epsilon \cos kx), \]  

(40)

where \( \epsilon \) is the amplitude and \( k \) is the “frequency” of the perturbation, results in a widely oscillating solution (Fig. 5b)

\[ n'(x) = \frac{1}{\sqrt{2\pi (w^2 - \sigma^2)}} e^{-x^2/(2(w^2 - \sigma^2))} \times \left[ 1 + \epsilon e^{-2(x^2/2w^2)} \cos \left( \frac{w^2}{w^2 - \sigma^2} kx \right) \right]. \]  

(41)

Obviously, for any \( \epsilon \), one can choose the perturbation frequency \( k \) so large that the perturbed equilibrium (41) will not remain non-negative everywhere. In this case, solution (41) no longer describes a biologically meaningful coexistence. Therefore, the coexistence of all populations along the trait axis can be destroyed by an arbitrarily small perturbation by conferring a relative advantage/disadvantage to very similar populations.

4. The general theory

4.1. Non-linear generalizations

In this section we demonstrate how the conclusions of the linear theory above generalize to the non-linear case. For this purpose, we define the two kinds of niches and the community matrix as derivatives of the (possibly non-linear) functions \( I, S \) and \( R \). We demonstrate that the volume measures depicted in Fig. 3 are still usable.
The impact niche vector is defined as the derivative of the impact map $I$

$$C_i = \frac{\partial I}{\partial r_i} \quad (j = 1, 2, \ldots, L). \quad (42)$$

It measures the impact of a small change of the population size. Obviously, this “differential” impact depends on the population sizes at which the derivative is taken. We suppress this dependence in the notation. Similarly, the (differential) sensitivity niche vectors are defined as

$$S_i = -\frac{\partial r_i}{\partial I} \quad (i = 1, 2, \ldots, L). \quad (43)$$

For finite-dimensional population regulation, derivatives (42) and (43) can be written component-wise as partial derivatives

$$C_{jk} = \frac{\partial I_k}{\partial r_j} \quad (k = 1, 2, \ldots, D; \ j = 1, 2, \ldots, L) \quad (44)$$

and

$$S_{ik} = -\frac{\partial r_i}{\partial I_k} \quad (k = 1, 2, \ldots, D; \ i = 1, 2, \ldots, L), \quad (45)$$

representing the two-way interaction between the $i$th population and the $k$th regulating variable. In a similar manner, if the environmental interaction is captured by a function, like in Section 3.4, then the two kinds of niche are specified by the functions $C_i(q)$ and $S_i(q)$ representing the local per capita impact on, and the sensitivity to, the regulating variable $I(q)$. Boxes 1 and 2 provide examples for calculating the niche vectors with pure resource competition and with joint regulation via competition and predation, respectively.

---

**Box 1. Impact and sensitivity in Tilman’s model.**

In a model of Tilman (1982), which is also the basis of the niche theory of Leibold (1995) and Chase and Leibold (2003), $L$ species compete for $D$ resources. In the most general version, the dynamics is specified as

$$\frac{1}{n_i} \frac{dn_i}{dt} = r_i = f_i(R_1, R_2, \ldots, R_D) - m_i$$

and

$$\frac{dR_j}{dt} = g_j(R_j) - \sum_{i=1}^{L} n_if_i(R_1, R_2, \ldots, R_D)H_{ij}(R_1, R_2, \ldots, R_D)$$

(Tilman, 1982, p. 14). Here, $R_j$ is the availability of resource $j$; $m_i$ is the mortality rate of species $i$; $f_i$ is the function describing the dependence of the growth rate of species $i$; and $H_{ij}$ is the function describing the amount of resource $j$ required to produce each new individual of species $i$. In the more specific version of the model,

$$g_j(R_j) = a(R_0^j - R_j),$$

where $a$ is a constant, and the complicated dependence for per-capita resource consumption $f_i(R_1, R_2, \ldots, R_D)H_{ij}(R_1, R_2, \ldots, R_D)$ is replaced by the constant $c_{ij}$ (Tilman, 1982, p. 76). Then, the resource dynamics becomes linearized as

$$\frac{dR_j}{dt} = a(R_0^j - R_j) - \sum_{i=1}^{L} n_ic_{ij}.$$ 

At fixed densities, the resource equilibrium point can be calculated as

$$R_j^* = R_0^j - \frac{1}{a} \sum_{i=1}^{L} n_ic_{ij}.$$ 

Then, the regulating variable, i.e., the resource depletion, is

$$I_j = R_0^j - R_j^* = \frac{1}{a} \sum_{i=1}^{L} n_ic_{ij} = \sum_{i=1}^{L} n_iC_{ij},$$

where components of the impact niche vector, as we defined, are

$$C_{ij} = \frac{c_{ij}}{a}.$$ 

Our sensitivity niche vector is

$$S_{ij} = -\frac{\partial r_i}{\partial I_j} = \frac{\partial f_i}{\partial R_j}.$$ 

This gradient vector is obviously perpendicular to the $r = \text{const.}$ isoclines, including the ZNGI, corresponding to $r = 0$. 

---

This text is an excerpt from the article "G. Meszéna et al. / Theoretical Population Biology 69 (2006) 68–87".
Here, we use the “keystone predator” model of Chase and Leibold (2003, p. 49) to demonstrate the applicability of our concepts in situations other than resource competition. Coexistence of two species (abundances: $N_1$ and $N_2$), that are regulated by a single resource $R$ and a single predator $P$, is investigated. The dynamics is specified as

$$\frac{dR}{dt} = c[S - R] - f_1 N_1 - f_2 N_2,$$

$$\frac{dN_i}{dt} = (f_i a_i R - m_i P - d_i) N_i, \quad i = 1, 2,$$

$$\frac{dP}{dt} = (m_1 c_1 N_1 + m_2 c_2 N_2 - d_P - 2P) P,$$

where the coefficients are self-explanatory. The vector of regulating variables is $I = (R, P)$. (As we have variables other than resources, we no longer restrict ourselves to the “exploitation” interpretation and sign convention.) The two regulating factors allow coexistence of at most two species on the trophic level of interest. We consider the predator as a part of the environment, instead of a species of our interest. We allow it to be regulated by factors other than dependence on “our” species. The term $-2P$ collects all the extra regulations. The equilibrium values of the regulating variables are

$$R^* = S - 1/c [f_1 N_1 + f_2 N_2],$$

$$P^* = 1/\alpha [m_1 c_1 N_1 + m_2 c_2 N_2 - d_P].$$

Both the impact and the sensitivity vectors have an “$R$” and a “$P$” component:

$$C_{ir} = \frac{\partial R^*}{\partial N_i} = -f_i/c, \quad C_{IP} = \frac{\partial P^*}{\partial N_i} = m_i c_i/\alpha,$$

$$S_{IR} = -\frac{\partial r_i}{\partial R} = -f_i a_i, \quad S_{IP} = -\frac{\partial r_i}{\partial P} = m_i.$$

This description breaks down for $\alpha = 0$, i.e., when the investigated populations are the only regulators of the predator. Then, there is no such thing as an equilibrium predator density at given values of $N_1$ and $N_2$, because the predator grows exponentially at fixed $N_i$. In this case, we have to include the predator in the list of the populations of interest and introduce new regulating variables to parametrize their interactions.

The (differential) community matrix is defined as the derivative of the population regulation $R$

$$a_{ij} = -\frac{\partial r_i}{\partial n_j} \quad (i, j = 1, 2, \ldots, L). \quad (46)$$

Again, its dependence on $n_j$ is suppressed in the notation. The element $a_{ij}$ measures the sensitivity of the growth rate of the $i$th population to the size of the $j$th population. The chain rule for derivatives establishes the same connection between the differential niche vectors and the differential community matrix as in the linear model

$$a_{ij} = S_{ij} \cdot C_j. \quad (47)$$

As known from multivariate calculus (Apostol, 1962, pp. 84–86; Edwards Jr., 1973, p. 245), the function $R$ maps the infinitesimal volume element $dn_1 dn_2 \ldots dn_L$ around the point $(n_1, n_2, \ldots, n_L)$ into a parallelepiped of volume $|J| dn_1 dn_2 \ldots dn_L$, where $J = \det a$ is the determinant of the community matrix evaluated at that point. That is, in complete analogy with the linear case, $|J|$ is the right measure of the local strength of population regulation on the community level. If $|J|$ is small, the growth rates depend weakly on the population sizes at least in some directions (Fig. 6). The impact function $I$ maps the volume element $dn_1 dn_2 \ldots dn_L$ into a parallelepiped of volume $V_C dn_1 dn_2 \ldots dn_L$. Accordingly, the volume $V_C$ is the measure of the strength of the environmental impact of the populations. Analogously, the volume $V_S$ measures the environmental sensitivity of the populations.

As a direct consequence of relation (47), Proposition of Section 3.3 remains valid. Accordingly, either a weak environmental impact, or a weak environmental dependence leads to a weak population regulation. With this connection between the community matrix and niche in place, a generalization of the linear result is established if we demonstrate the detrimental effect of small $|J|$ on the robustness of coexistence.

An explicit differentiation of the set of equilibrium equations

$$r_i(E, I(n_1, n_2, \ldots, n_L)) = 0 \quad (i = 1, 2, \ldots, L) \quad (48)$$
is represented by the excess mortality $\Delta_i$ ($i = 1, 2, \ldots, L$).

Accordingly, we write the dependence on the external variables in the form

$$r_i(E, I(n_1, n_2, \ldots, n_L)) = \hat{r}_i(I(n_1, n_2, \ldots, n_L)) - \Delta_i(E).$$

(Negative $\Delta_i$ represents decreased mortality or increased birth rate.) Then the equilibrium equations become

$$\hat{r}_i(I(n_1, n_2, \ldots, n_L)) = \Delta_i(E) \quad (i = 1, 2, \ldots, L).$$

We collect the variables $n_i$, $\hat{r}_i$ and $\Delta_i$ in the vectors $\mathbf{n}$, $\hat{\mathbf{r}}$ and $\Delta$. The probability distribution of the environmental parameters translates into a (supposedly continuous) probability distribution of the excess mortalities, represented by the probability density $p(\Delta)$. Then the probability $P(\mathbf{n} > 0)$ of having all equilibrium densities positive satisfies

$$P(\mathbf{n} > 0) = \int_{n > 0} \frac{p(\hat{\mathbf{r}}(\mathbf{n}))}{k[\hat{\mathbf{r}}(\mathbf{n})]} |J(\mathbf{n})| \, d\mathbf{n},$$

where $k[\mathbf{r}]$ denotes the number of all positive $n$ values for which $q = \hat{\mathbf{r}}(\mathbf{n})$. (See Stromberg (1981, p. 385) for $k = 1$ and Rado and Reichelderfer (1955, p. 363) for the general case.)

It is clear from Eq. (52) that the probability that all $L$ species coexist goes to zero when $|J(\mathbf{n})| \to 0$ for all relevant $\mathbf{n}$.

We stress that this probabilistic argument does not assume any stochasticity in the behavior of the populations. We calculated the probability of finding the proper conditions for coexistence in a randomly chosen location. A community with small $|J|$ has a small chance of finding the permitting combinations of the external parameters.

Our findings can be summarized as follows:

**Limiting Similarity Principle.** Any of the following conditions imply the next one and make coexistence of a given set of populations improbable, i.e., restricted to a narrow range of the external environmental parameters:

- Large overlap between either the impact or the sensitivity niches.
- Small $\mathcal{V}_C$ or $\mathcal{V}_S$.
- Small $|J|$.

The **Discrete Competitive Exclusion Principle** is a limiting case. If its condition $L \leq D$ is violated then $L$ number of $D$-dimensional vectors cannot be linearly independent, implying $\mathcal{V}_C = \mathcal{V}_S = |J| = 0$ and structural instability of coexistence.

5. **Discussion**

5.1. **Competitive exclusion and limiting similarity**

The “competitive exclusion principle” is a controversial, but central, reference point in community ecology. By the “admittedly unclear” words of Hardin (1960), the principle is stated as “complete competitors cannot coexist”; or, “ecological differentiation is the necessary condition for
coexistence”. We removed the ambiguity by a mathematical characterization of the nature and extent of the required differentiation. “Limiting similarity” is a robust prediction in a well-defined sense. While there is no absolute lower bound for the permitted similarity, coexistence of similar populations is possible only for a narrow range of external parameters. Although the robustness/probability of coexistence of a given set of populations can be calculated only in a specific model context, the idea of shrinking the possibility of coexistence with increasing similarity is general.

May and MacArthur (1972) and May (1973, 1974) concluded that a strong enough population regulation, therefore limiting similarity, was needed to compensate for the effect of noise and to prevent the extinction of some of the populations. While this conclusion is fully consistent with our analysis, Rosenzweig (1995, p. 127) interpreted May’s results as an early recognition of the lack of limiting similarity in a constant environment. We emphasize that limiting similarity is the expectation even in a constant environment.

As limiting similarity is the general expectation, it is of no surprise that the continuous coexistence in Roughgarden’s (1979) example is structurally unstable, as we demonstrated in Section 3.5. Lack of structural stability of Roughgarden’s model was already observed implicitly by Sasaki and Ellner (1995), Sasaki (1997) when they investigated unrelated but mathematically isomorphic models. (We thank Yoh Iwasa for pointing out this connection. See also Haccou and Iwasa (1995, 1998).) Szabó and Meszéna (in press) analyzed Roughgarden’s model beyond the scope of the small perturbations studied here. They demonstrated the generic validity of the usual “limiting similarity” expectations with respect to the Lotka–Volterra model. The possibility, but exceptionality, of continuous coexistence was also noted by Geritz (1995) and Geritz et al. (1999) for asymmetric seed competition, by Kitzing et al. (1999) in a metapopulation context and by Meszéna and Szathmáry (2001) for prebiotic replicators. Gyllenberg and Meszéna (2005) reviewed the issue and proved a very general statement: Coexistence of infinitely many populations in a finite volume of the phenotype space is structurally unstable for any model constructed from smooth ingredients.

5.2. Connection to the dynamics of adaptation

Ecological conditions for coexistence are related to frequency-dependent selection (Christiansen, 1988; Heino et al., 1998). Accordingly, both the general impossibility and the exceptional possibility of coexistence of similar strategies are of evolutionary relevance. A general theory of frequency-dependent selection (Metz et al., 1996a; Geritz et al., 1997, 1998; Meszéna et al., 2005) demonstrates that similar phenotypes can coexist only at the “singular” points of the phenotype space, where the selection gradient disappears. Competitive exclusion, i.e., the “survival of the fittest”, operates away from the singular points in line with the usual picture of natural selection (see also Dieckmann and Law (1996), for directional evolution in this context). On the other hand, the possibility of coexistence in the vicinity of singular points allows “branching”-type evolution, which was suggested to be related to speciation (Dieckmann and Doebeli, 1999; Doebeli and Dieckmann, 2003; Kisdi and Geritz, 1999; Geritz and Kisdi, 2000; Meszéna et al., 1997; Mizera and Meszéna, 2003; Schreiber and Tobiason, 2003; Geritz et al., 2004). This way, although coexisting species generically must differ considerably, and live in different niches, a new species can emerge via small modifications. (See Meszéna and Metz (1999) for the specific connection between the current theory, frequency-dependent selection and speciation.)

5.3. Population regulation and coexistence

Regulation of a single population and coexistence of several populations are usually discussed separately. We offered a unifying viewpoint. Robust coexistence, i.e., the lack of competitive exclusion, is equivalent to the existence of separate population regulation for each population, i.e., significant differences both in the way in which the populations affect the regulating variables and in the way in which the population growth rates depend on these variables. Without such differentiation, competitive exclusion, i.e., the impossibility of coexistence, is the default relation between two populations.

Competitive exclusion is a relatively fast process, except when $r_1 \approx r_2$. (Here, $r_1$ and $r_2$ denote the growth rates of the two species involved.) According to Eq. (5), the time-scale of the competitive exclusion is $1/(r_1 - r_2)$. For instance, a permanent 1% difference either in the birth or in the death rates results in competitive exclusion on the approximate time-scale of 100 generations.

Equality of the demographic parameters of two different species, without a specific reason, is highly unlikely. Either the two species have to be truly identical in all ecological and demographic respects, or the equality of the growth rates has to be established by a regulating feedback. As examples for the first case, sibling species (Turner, 1999) may be able to coexist without niche segregation. (Loss of one of the two identical species via demographic stochasticity can be extremely slow when the population sizes are large enough.) However, as soon as discernible differences—other than in their mating traits—arise between the species, coexistence requires that their fitness difference is made zero by differential regulation (Galis and Metz, 1998). In this paper, we are concerned with the latter type of regulated (or stable, Chesson, 2000b) coexistence. See Goldschmidt et al. (1990), Seehausen and Bouton (1997), Bouton et al. (1997) for detailed investigations of niche segregation, i.e., the emergence of differential regulation, between recently formed Cichlid species.
Chesson (1994, 2000a,b) investigated the coexistence-stabilizing mechanisms by comparing the growth rates of the populations, when rare, and when they are in equilibrium: rarity must provide an advantage. In contrast, our approach tests the effect of small abundance changes on the growth rates. The extra complication of our differential approach pays off in the linearity of the perturbation and in allowing to reach the limiting similarity conclusion. Actually, our results rely on a condition weaker than dynamical stability. Structural stability (robustness) requires det $a$ to be non-zero (large enough in absolute value), while linear dynamical stability is based on the stronger condition of a negative real part of all eigenvalues of the population matrix $a$.

By now, a consensus has been reached on the traditionally “controversial” (Krebs, 2001, p. 290) issue of population regulation. It has been established that any population has to be regulated (Turchin, 1999, 2003; Hanski, 1999). However, as the notion of regulating factors is not recognized in community ecology in its generality, controversies about competitive exclusion persist (cf. den Boer, 1986; see Gorskov et al. (2000) for a consistent picture, different from the ours). It is a general view that competitive exclusion and niche are relevant issues only for communities shaped by resource competition (Begon et al., 1996, p. 265). It is a widely accepted argument that “…there are many situations, where disturbance, such as predation, storms on a rocky shore or frequent fires, may hold down the densities of populations, so that the resources are not limiting and the individuals do not compete for them.” (Begon et al., 1996, p. 802). Then, it is assumed that reduced competition reduces the danger of competitive exclusion and increases diversity beyond what is permitted by “competition” (Huston, 1979; Begon et al., 1996, p. 802).

This type of reasoning is misleading. There is no essential difference between resource competition and the other types of population regulation, like predation and disease spread (cf. Case and Leibold, 2003). Resources are not limiting/regulating only when other factors regulate. Then, the structure of that regulation has to be studied in order to predict the effect on diversity. For instance, in the simplest case, a single predator contributes a single regulating factor, so it can allow one, and only one, additional species to coexist (see Box 2). In particular, the suggestion that “disturbance” slows down competitive exclusion (see Huston, 1979, 1994) is baseless, as disturbance does not habitually decrease the difference between the growth rates.

Because of the propensity of the regulating factors to be high in number, we need a method to count only the really different ones. In line with Abrams (1988), our version of Limiting Similarity Principle in Section 4 provides the necessary background: the limited overlap of the niches, the impact ones as well as the sensitivity ones is the criterion (see also Meszéna and Metz, 1999). In this way we unified the discrete and the continuous versions of the theory of competitive exclusion.

5.4. Spatially extended populations

Beyond being fully identical, an extremely strong migration barrier between the populations is the only mathematical possibility of avoiding the necessity of regulated coexistence: if the migration between two habitats is negligible even on the evolutionary time-scale then there is no danger of competitive exclusion between the species living separately.

Else, the populations equivlate between the habitats on the evolutionary time-scale. Then, our theory applies if we identify the $r_i$s with the overall population growth rates after the spatial relaxation. The overall rates are the averages of the local ones, weighted by the relaxed local abundances (Chesson, 2000a).

The theory of structured populations (Metz and Diekmann, 1986; Caswell, 2000; Diekmann et al., 1998, 2001, 2003) applies for calculating the $r_i$s as well as the equilibrium distributions. In this paper, we are interested in either the equilibrium of the populations or small perturbations of this equilibrium. Then the analysis of Greiner et al. (1994) establishes the time-scale separation between the slow overall growth/decline of the population and the faster spatial relaxation. (Here we assume a finite world with a bounded migration time-scale.) Populations, which are structured in any other respect, can be studied similarly.

When a spatially extended population is considered, the overall $E$ and $I$ vectors should include the values of the external parameters and the regulating variables at each location, respectively (cf. the variables $E_{js}$ and $C_{js}$ of Chesson, 2000a). For instance, in the two-patch case of Meszéna et al. (1997) the population is regulated by a single factor in both patches separately. Then, $I$ is the two-dimensional vector of the two local regulating variables. The analogous situation for an environmental gradient is studied by Czarán (1989), Mizera and Meszéna (2003), Mizera et al. (in prep). Here the regulation is described by a continuum of regulating variables, i.e., by a location-dependent regulator. An important conclusion of these studies is that the environmental tolerance and the mobility of the competitors together determine the extent of the region occupied by a single species.

In general, even a single resource can act as a continuum of regulating factors along an environmental gradient. The environmental condition at the location of the resource should be considered as a quality of that resource. Examples include the light gradient (Hutchinson, 1961) or the temperature gradient (Tilman et al., 1981) in a water volume, or the feeding space separation of warblers (MacArthur, 1958).

5.5. Fluctuating environment

Spatio-temporal fluctuations lead to increased diversity easily. As Armstrong and McGehee (1976); Hsu et al. (1977) demonstrated, it is possible that two species feeding

Still, as a non-regulated population fluctuates around a path of exponential growth/decline (Tuljapurkar, 1990; Metz et al., 1992; Ferrière and Gatto, 1995), spatio-temporal heterogeneity does not weaken the necessity of having regulating/stabilizing mechanisms to equalize the long-term growth rates (Chesson, 1991; Hanski, 1999, p. 29). While competitive exclusion may be irrelevant locally, if a species is going to survive longer than the life-span of a single habitat, their regulation and competitive exclusion have to be considered globally, on a metapopulation level (Tilman et al., 1994; Parvinen and Meszéna, in prep.).

Chesson (1994, 2000b) distinguish between two types of fluctuation-induced stabilizing mechanisms: “storage effect” and “effect of relative non-linearity”. The first one essentially corresponds to temporal niche segregation allowed by the fluctuation (cf. Christiansen and Fenchel, 1977, p. 69), while the second one means that the descriptors of the fluctuation (like variance) emerge as additional regulating variables via the non-triviality of non-linear averaging (cf. Levins, 1979; Kisdi and Meszéna, 1993).

These results make it extremely plausible that our picture of decreasing robustness with increasing similarity can also be extended to fluctuation-mediated coexistence in stationarily fluctuating environments. However, this is beyond the scope of the current paper.

5.6. Niche space as a map for competitive exclusion

According to Hutchinson’s (1978, pp. 152–157) historical analysis, the notion of a niche acquired its connection to Gause’s principle gradually. While the relation between the niche overlap and the strength of competition is considered as the “central tenet of niche theory” (Case, 2000, p. 368), the actual usage of the term is not always consistent with this tenet (Juhasz-Nagy, pers. comm.). For consistency, the niche space has to be equated with the set of regulating factors/variables. This set can be either a finite/discrete one, or a continuum. In the latter case, the niche space is a Euclidean space of the “niche dimensions”, which parametrize the continuum of the regulating factors (e.g., the resource quality $q$ in Sections 3.4 and 3.5). The niche of a species is given by specifying its impact on, and sensitivity towards, the regulating factors, a direct generalization of the notion of the resource utilization function. Such a description of the niche of a species consists of two finite-dimensional vectors, or two functions, depending on the finite or continuous nature of the niche space. We established that a small overlap between both the impact and the sensitivity niche vectors/functions, i.e., the proper separation in the niche space, is the general prerequisite for robust coexistence.

Note that the niche space is not a pre-existing emptiness, which is filled out by the species later. Any new species may extend the niche space by contributing additional regulating factor(s).

Hutchinson’s imagination was captured by the continuous case. He distinguished between two types of niche dimensions: resource quality (“bionomic”) and environmental condition (“scenopoetic”) (Hutchinson, 1978, p. 171). The Gauseian idea behind this distinction is clear. Reduced competition between two species can arise either by differing in local resource use or by living separately; the latter possibility is related to the different requirements of the species with respect to the ecological conditions. The connection between niche segregation and reduced competition is direct in the first case, but mediated via spatial distributions in the second one. Not accidentally, efforts of modeling niche segregation usually refer to a niche dimension of resource quality type (MacArthur and Levins, 1967; Yodzis, 1989, p. 119). Our picture unifies these two kinds of niche dimensions by recognizing that environmental conditions can also be seen as a quality of the resource (Section 5.4). (See also Christiansen and Fenchel, 1977, p. 69.)

While, undoubtedly, Gause’s principle had a central role in Hutchinson’s thinking, his description of the niche as the “hypervolume” in the space of “all variables that affect a particular species” was too vague to secure the connection. Following this definition, environmental conditions are often regarded as typical niche variables without any reference to population regulation. It was even suggested that considering resource quality as a niche dimension had been a logical mistake made by Hutchinson. According to this view, the abundance, instead of the quality, of the food is the “right” niche dimension, as this is the “condition”-type quantity (Leibold, 1995). However, the central tenet is lost in this way. Living under the same environmental conditions does not imply competitive exclusion, resource abundances counted or not. Note that resource partitioning is a partitioning of the space of the resource qualities and not of the space of resource abundances. (On the other hand, the supply rate of a resource is a proper “condition”-type niche dimension; Köszeghy, 2004.) See also Begon et al. (1996), who introduces the notion of niche with the pure “condition” interpretation on p. 87, but stress that the concept is “most powerful” in a resource context, while they remain undecided about the quality-or-abundance issue on p. 130.

Hutchinson’s distinction (p. 161) between the fundamental and the realized niche can be represented in our framework naturally by evaluating the defining derivatives (42) and (43) in the absence and in the presence of the competitors, respectively.
The possibility of a niche “space” consisting of a finite number of points seems to contradict radically the “hypervolume” notion of Hutchinson. However, from a biological, as opposed to a formal, point of view, this is the right picture for the discrete case. If the populations are limited by a finite number of resources, partitioning of this finite set of resources is the proper analogy to the partitioning of the resource quality continuum. The discrete notion of niche, as a finite-dimensional vector, was recognized already by Petraitis (1989).

The discrete version of the resource competition theory was rejuvenated by Tilman (1982). He emphasized the need for “mechanistic” modeling, i.e., for an explicit representation of the feedback loop. Leibold (1995), Chase and Leibold (2003) re-interpreted the notion of niche in the context of Tilman’s model. Their approach has strong similarities to the one which was pursued in this paper:

- They stress the importance of regulating factors other than resources, especially the regulating role of predators.
- They stress the importance of considering the two legs of regulation: the population’s impact and dependence on the regulating factors.

Our notion of the impact niche, i.e., the per capita consumption rate c on the regulating factors, is very similar to Leibold’s one, i.e., the per capita consumption rate c. (See Box 1 for the explicit connection between the two. The slight difference was motivated by the generality of our choice.) Leibold’s “requirement” niche of a species is the region of the hyperspace of the abundances of the finite number of resources (or other regulating factors). Instead, we proposed “sensitivity” niche vectors in a role very symmetrical to the impact niche. It is clear that, in our representation, the two components of the niches have equal roles in determining both the existence and the stability of the population dynamical fixed point.

5.7. Outlook

Note that the main question of this paper, i.e., that “What is the probability of the possibility of coexistence of a predefined set of species?” is not identical to the empirically more relevant question of “What is the probability distribution of similarity of coexisting species in nature?”. The answer to the second question is affected by the distribution of the available species as well as the assembly and evolutionary processes, (e.g., Leibold, 1998) which are beyond our discussion.

Nevertheless, we may formulate general expectations based on the presented results. We expect divergence in species specific traits involved in the physiological, behavioral mechanisms determining the impact and sensitivity niches of coexisting species, i.e., in their relations towards the regulating factors. Studies on character displacement (Schluter, 2000a,b, p. 296) demonstrate this empirically in a very convincing way. On the other hand, coexistence is obviously facilitated by similar requirements with respect to the external parameters, i.e., in traits not related to population regulation. (We thank Mathew Leibold for pointing out to this dichotomy.) When similarity measures are based on phylogenetic relationships (like in Webb (2000)), these two kinds of similarity are mixed, and the results are indeterminate. We think that the mechanistic approach, that makes explicit the way in which populations are regulated, is needed for community studies to avoid such confusions.

In this paper, we concentrated on the connection between niche overlap and competitive exclusion. This theory should be complemented by analyses on different levels.

On the lower level, one has to understand how the shape and the breadth of the niche of a species is determined by the physiological/behavioral constraints (i.e., by the trade-offs) and/or by the evolutionary factors. (See also Chesson (2000b) for the role of trade-offs in coexistence.) Schreiber and Tobiason (2003) provide an example. They investigated generalist/specialist evolution in the context of the 2-resource Tilman model. Tilman’s distinction between essential and substitutable resources—which is related to the convex-or-concave nature of the appropriate trade-off (cf. Levins, 1962) but played no role in our discussions because of the linearization— is essential in this analysis. A behavioral constraint (i.e., the fixed level of mobility) is important in determining the spatial niche-breadth along an environmental gradient in Mizera and Meszéna (2003), Mizera et al. (in prep.).

On a higher level, connection to the wider problem of biodiversity is less clear. As an ecosystem is not necessarily saturated, the actual diversity may be determined by the balance between colonization and extinction, as suggested by MacArthur and Wilson (1967). According to Hubbel (2001), the relation between the niche-assembly and dispersal-assembly perspectives is “… one of the most fundamental unsolved problems in ecology today”. Here we attempted to clarify the first one of these two pictures, which is a pre-requisite for reconciliation.

Acknowledgments

We thank Sándor Bartha, Niels Bouton, Tamás Czárán, Yoh Iwasa, Patsy Haccou, Éva Kisdi, Kolos Kőszeghy, Michel Loreau, István Scheuring, Akira Sasaki, Beáta Oborny and the late Pál Juhasz-Nagy for discussions. Comments by Peter Chesson, Mathew Leibold and an anonymous referee on the first version of the ms were helpful. This work was financed from the Grants OTKA T033097, T049689 and NWO-OTKA 048.011.039.

Appendix A

Section A introduces the basic notions in more precise terms. Section B attempts to de-mystify the concept of tensors in this context, which will be needed to calculate the volume of the parallelepiped in Section C.
A.1. Unifying the finite- and infinite-dimensional cases: the basic notions

Let \( \Omega \) denote the set of the (descriptors of the) regulating factors. \( D \) is the number of these factors. When \( D = \infty \), \( \Omega \) is usually a subset of \( \mathbb{R}^n \). The choice \( \Omega = \mathbb{R} \) corresponds to the one-dimensional resource continuum of Section 3.4.

The state of the regulating environment, denoted by \( I \), is specified when the value of each regulating variable is given. Consequently, \( I \) can be considered as a mapping from \( \Omega \) to \( \mathbb{R} \). By using the usual notation \( \mathbb{R}^D \) for the set of all functions from \( \Omega \) to \( \mathbb{R} \),

\[
I \in \mathbb{R}^D : \quad x \in \Omega \rightarrow I(x) \in \mathbb{R}.
\]  

(A.1)

When \( D \) is finite, \( \Omega \) is identified with the index set \( \{1, 2, \ldots, D\} \) and \( \mathbb{R}^D \) becomes \( \mathbb{R}^n \). In this case, \( I \) corresponds to a finite-dimensional vector:

\[
I = (I_1, I_2, \ldots, I_D) \in \mathbb{R}^n.
\]  

(A.2)

Analogously, the impact and the sensitivity vectors/functions \( C \) and \( S \) for the population \( j \) are elements of \( \mathbb{R}^D \), or \( \mathbb{R}^D \) when \( D \) is finite. Both assign a value to each factor \( f \).

The scalar product for these vectors/functions in \( \mathbb{R}^D \), measuring their overlap, is defined in the usual way

\[
u \cdot v = \sum_{i=1}^{D} u_i v_i, \quad \Omega \text{ is finite,}
\]  

(A.3)

\[
u = \sqrt{\nu \cdot \nu}.
\]  

(A.4)

A mathematical caveat is needed here though: for the infinite-dimensional case, the existence of the “overlap” integral in (A.3) is an additional requirement, which restricts the allowed functions \( I \) to the so-called \( L_2 \) space.

In some applications, the set \( \Omega \) can be more complicated than \( \mathbb{R}^n \). Always, a prescription for the scalar product is needed, which makes the allowable subset of \( \mathbb{R}^D \) a Hilbert space (Rudin, 1974, p. 76). For \( D = \infty \), definition (43) of the sensitivity vector/function and the chain rule (47) relies on the Riesz representation theorem for Hilbert spaces (Theorem 4.12, in Rudin, 1974, p. 85).

Appendix B. Tensors light

While tensor calculus is usually considered difficult, the parts that we need are very simple. From our point of view, an \( L \)-tensor is a function with \( L \) arguments in \( \Omega \).

The tensor product, denoted by \( \odot \), of \( L \) functions \( u_1, u_2, \ldots, u_L \in \mathbb{R}^D \) is an \( L \) tensor:

\[
(u_1 \odot u_2 \odot \cdots \odot u_L)(x_1, x_2, \ldots, x_L) = u_1(x_1)u_2(x_2) \cdots u_L(x_L).
\]  

(B.1)

When \( \Omega \) is finite, the tensor product is a quantity with \( L \) indices, i.e., an “\( L \)-dimensional matrix”:

\[
(u_1 \odot u_2 \odot \cdots \odot u_L)_{j_1j_2 \cdots j_L} = u_{j_1}u_{j_2} \cdots u_{j_L}
\]  

(B.2)

where \( u_{j_l} \) denotes the \( j_l \)th component of the vector \( u_l \). It appears intimidating, but no deep issues are involved here. Note that the tensor product is identical to the usual dyadic product of matrix calculus when \( L = 2 \). In this case, the result is a two-dimensional matrix.

The scalar product and norm of tensors are defined analogously to Eqs. (A.3) and (A.4)) by extending the integration/summation to all variables/indices. In particular, the scalar product of two tensor products can be evaluated componentwise, as

\[
(u_1 \odot u_2 \odot \cdots \odot u_L) \cdot (v_1 \odot v_2 \odot \cdots \odot v_L) = (u_1 \cdot v_1)(u_2 \cdot v_2) \cdots (u_L \cdot v_L).
\]  

(B.3)

Appendix C. Volume of parallelepiped

The \( (L \)-dimensional) volume \( V_u \) of the parallelepiped, spanned by the \( L \) vectors \( u_1, u_2, \ldots, u_L \), has a central role in our analysis. The issue is usually discussed in the context of differential geometry and general relativity (see, for instance, Edwards Jr., 1973, p. 345; Mishra et al., 1973, p. 204). For the convenience of the reader, we summarize the necessary mathematics here.

The case \( L = D \) is well-known (Apostol, 1962, pp. 84–86; Case, 2000, pp. 420–422; Edwards Jr., 1973, p. 245). The volume can be calculated as the determinant of the square matrix \( u \) formed from the vectors, i.e.,

\[
V_u = \det u = \sum_{\sigma \in \Sigma_L} (-1)^{\sigma} u_{\sigma(1)} \cdot u_{\sigma(2)} \cdots u_{\sigma(L)}
\]  

(C.1)

for \( L = D \).

Here, the generic element \( u_{\sigma(j)} \) of the matrix \( u \) is the \( j \)th component of the \( \sigma \)th vector \( u_\sigma \). In the expansion of the determinant, \( \Sigma_L \) denotes the group of the permutations of the index set \( \{1, \ldots, L\} \) and \(( -1)^\sigma \) is \( \pm 1 \) if the permutation \( \sigma \) is even or odd. We have to generalize this formula to \( L \neq D \), when the matrix \( u \) is not a square matrix and the determinant has no sense. (\( D \) is allowed to be infinite, \( L \) is not.)

We will show that the volume of the parallelepiped spanned by \( L \) vectors \( u_1, u_2, \ldots, u_L \) can be written as

\[
V_u = |u_1 \wedge u_2 \wedge \cdots \wedge u_L|,
\]  

(C.2)

i.e., it is the norm of the so-called wedge product of the vectors. The wedge product is defined as the \( L \)-tensor

\[
(u_1 \wedge u_2 \wedge \cdots \wedge u_L) = \frac{1}{\sqrt{L!}} \sum_{\sigma \in \Sigma_L} (-1)^{\sigma} u_{\sigma(1)} \odot u_{\sigma(2)} \odot \cdots \odot u_{\sigma(L)}.
\]  

(C.3)

That is, \( \wedge \) can be interpreted as the unique antisymmetrized tensor product of the vectors. The prefactor
$1/\sqrt{L!}$ is chosen for our later convenience. Observe the similarity to the definition of the determinant.

To develop some intuition for the wedge product first consider a special case. For $L = 2$, the wedge product can be written as

$$u \wedge v = \frac{1}{\sqrt{2}}(u \circ v - v \circ u), \quad \text{(C.4)}$$

or, for finite $D$, as

$$(u \wedge v)_j = \frac{1}{\sqrt{2}}(u_i v_j - u_j v_i) \quad \text{(C.5)}$$

which is an $L \times L$ antisymmetric matrix. In the special case of $D = 3$, the antisymmetric matrix (C.5) has 6 non-zero components, which are, after multiplication with $\pm 1/\sqrt{2}$, identical to the 3 components of the vector product $u \times v$. (See Anton, 1984, p. 111 for the vector product.) Then, it is clear that $|u \wedge v| = |u \times v| = |u||v| \sin(u, v). \quad \text{(C.6)}$

In line with our claim, this is just the area of the parallelogram spanned by the vectors $u$ and $v$ [cf. Eq. (22)]. The wedge product can be seen as a generalization of the vectorial product, which itself is a peculiarity of the 3-dimensional spaces.

When the $D$ is finite, one can spell the general wedge product (C.3) out in index notation, as

$$(u_1 \wedge u_2 \wedge \cdots \wedge u_L)_{j_1 j_2 \cdots j_L} = \frac{1}{\sqrt{L!}} \sum_{\sigma \in S_L} (-1)^\sigma u_{\sigma(1)}_{j_1} \cdot u_{\sigma(2)}_{j_2} \cdots \cdot u_{\sigma(L)}_{j_L}. \quad \text{(C.7)}$$

In this $L$-tensor, the only elements that can be non-zero are those where the indices $j_1, j_2, \ldots, j_L$ are all different. To prove (C.2), we discuss three cases.

For $D < L$, the wedge product disappears, as there are no $L$ different indices. This is in line with the fact that the $L$-dimensional volume $V_u$ is zero in $D < L$ dimensions.

For $D = L$, each non-zero element equals $\pm \det u/\sqrt{L!}$, where the square matrix $u$ is constructed from the vectors $u_i$. As we have $L!$ of them, the norm of the wedge product is

$$|u_1 \wedge u_2 \wedge \cdots \wedge u_L| = |\det u| = V_u \quad \text{for } L = D. \quad \text{(C.8)}$$

For $D > L$, $u$ is no longer a square matrix and $|\det u|$ has no meaning. Nevertheless, expression (C.2) remains valid, as one can apply the argument, leading to Eq. (C.8), in the $L$-dimensional subspace of $\mathbb{R}^D$ containing the vectors $u_i$. This proves our claim.

The Proposition of Section 3.3 is a direct consequence of the relation

$$(S_1 \wedge S_2 \wedge \cdots \wedge S_L)(C_1 \wedge C_1 \wedge \cdots \wedge C_L) = \frac{1}{\sqrt{L!}} \sum_{\sigma \in S_L} (-1)^\sigma S_{\sigma(1)} \circ \cdots \circ S_{\sigma(L)}$$

$$\times \left(\frac{1}{\sqrt{L!}} \sum_{\sigma' \in S_L} (-1)^{\sigma'} C_{\sigma'(1)} \circ \cdots \circ C_{\sigma'(L)}\right)$$

$$= \frac{1}{L!} \sum_{\sigma \in S_L} \sum_{\sigma' \in S_L} (-1)^{\sigma+\sigma'} (S_{\sigma(1)} \cdot C_{\sigma'(1)})$$

$$\times (S_{\sigma(2)} \cdot C_{\sigma'(2)}) \cdots (S_{\sigma(L)} \cdot C_{\sigma'(L)})$$

$$= \sum_{\sigma \in S_L} (-1)^\sigma (S_1 \cdot C_{\sigma(1)})$$

$$\times (S_2 \cdot C_{\sigma(2)}) \cdots (S_L \cdot C_{\sigma(L)}) = \mathbf{a}, \quad \text{(C.9)}$$

where the matrix $a$ is defined by $a_{ij} = S_i \cdot C_j$. \quad \text{(C.10)}$

in line with Eqs. (12) and (47).

Finally, we must demonstrate that

$$(u \wedge v)^2 + (u \cdot v)^2 = u^2 \cdot v^2, \quad \text{(C.11)}$$

which is the basis of Eq. (25). Consider that

$$(u \wedge v)^2 = \frac{1}{2}(u \circ v - v \circ u)^2$$

$$= (u \circ v)^2 - [(u \circ v)(v \circ u)]. \quad \text{(C.12)}$$

Then, the equalities

$$(u \circ v)^2 = u^2 \cdot v^2 \quad \text{(C.13)}$$

and

$$(u \circ v)(v \circ u) = (u \cdot v)^2 \quad \text{(C.14)}$$

prove the statement.

As an illustration we evaluate the area of the parallelogram for the resource continuum case of Section 3.4, i.e., for $\Omega = \mathbb{R}$, $L = 2$

$$V_u = |u_1 \wedge u_2| = \sqrt{\int_{x_1} \int_{x_2} \left[\frac{u_1(x_1)u_2(x_2) - u_1(x_2)u_2(x_1)}{\sqrt{2}}\right]^2 dx_1 dx_2. \quad \text{(C.15)}$$

A straightforward calculation connects this expression to the overlap integral:

$$V_u = \sqrt{\int_x [u_1(x)]^2 dx \cdot \int_x [u_2(x)]^2 dx} - \left[\int_x u_1(x)u_2(x) dx\right]^2. \quad \text{(C.16)}$$

This is in line with formulas (25) and (C.11): the area decreases with increasing overlap.


Szabó, P., Meszéna, G., Limiting similarity existing. Oikos, in press.


