

Stability of the replicator equation for a single species with a multi-dimensional continuous trait space

Ross Cressman^{a,*}, Josef Hofbauer^b, Frank Riedel^c

^a*Department of Mathematics, Wilfrid Laurier University, Waterloo, Ontario, N2L 3C5, Canada*

^b*Department of Mathematics, University College London, London WC1E 6BT, UK*

^c*Department of Economics, University of Bonn, Adenauerallee 24, 53113 Bonn, Germany*

Received 13 April 2005; received in revised form 27 July 2005; accepted 28 July 2005

Available online 24 October 2005

Dedicated to the memory of John Maynard Smith

Abstract

The replicator equation model for the evolution of individual behaviors in a single species with a multi-dimensional continuous trait space is developed as a dynamics on the set of probability measures. Stability of monomorphisms in this model using the weak topology is compared to more traditional methods of adaptive dynamics. For quadratic fitness functions and initial normal trait distributions, it is shown that the multi-dimensional continuously stable strategy (CSS) of adaptive dynamics is often relevant for predicting stability of the measure-theoretic model but may be too strong in general. For general fitness functions and trait distributions, the CSS is related to dominance solvability which can be used to characterize local stability for a large class of trait distributions that have no gaps in their supports whereas the stronger neighborhood invader strategy (NIS) concept is needed if the supports are arbitrary.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Adaptive dynamics; CSS; NIS; Replicator equation; Neighborhood superiority; Strategy dominance; Measure dynamics; Weak topology

1. Introduction

Dynamical systems on the set of probability measures over a continuous trait space have been developed as one means to predict the evolution and stability of distributions of individual behaviors in a biological species (Bomze, 1990, 1991; Oechssler and Riedel, 2001, 2002). These systems generalize the well-known replicator equation approach of dynamic evolutionary game theory (Hofbauer and Sigmund, 1998; Cressman, 2003) when the trait space is finite (i.e. when there are a finite number of pure strategies) and individuals interact in random pairwise

encounters that determine their payoffs (i.e. their fitness or reproductive success). A primary objective of these theoretical models is the characterization, in terms of static payoff/fitness comparisons, of static conditions (and distributions that satisfy the conditions) that imply dynamic stability. Such conditions then allow practitioners of evolutionary game theory to describe the outcome of the evolutionary process without a detailed analysis of the underlying dynamical system.

For instance, when there are n pure strategies (so trait space is finite) and there are random pairwise encounters that contribute additively to fitness, the static equilibrium concept of an evolutionarily stable strategy defined by Maynard Smith (1982) (i.e. a strategy for which, whenever all individuals use this strategy, the population cannot be invaded by a rare mutant under the influence of natural selection) has been a huge success since the resultant heuristic static conditions have a clear biological basis. We will refer to this case as the matrix game model (Meszéna et al., 2001) and the equilibrium concept as a matrix-ESS (Vincent and Cressman, 2000; Cressman and Hofbauer,

*Corresponding author. Tel.: +519 884 1970; fax: +529 884 9738.

E-mail addresses: rcressma@wlu.ca (R. Cressman),

j.hofbauer@ucl.ac.uk (J. Hofbauer), friedel@uni-bonn.de (F. Riedel).

¹This research was initiated while R.C. was a Fellow at the Collegium Budapest. Partial support from the German Science Foundation (Transregio “Governance and the Efficiency of Economic Systems”), from UCL’s Centre for Economic Learning and Social Evolution (ELSE) and from a NSERC of Canada Individual Discovery Grant is also acknowledged.

2005) since payoffs are given through an $n \times n$ payoff matrix.²

For continuous trait spaces, an alternative means to predict the evolutionary outcome is the *adaptive dynamics* method that has generated an enormous literature (see Abrams, 2001 and the references therein) since the phrase was introduced by Hofbauer and Sigmund (1990). This method is especially useful when the resident biological species is monomorphic (i.e. when all individuals in the population are using the same strategy) and there is a one-dimensional continuous trait space. Here, adaptive dynamics predicts stability of a monomorphic equilibrium if, for all other monomorphisms that are small perturbations of this equilibrium, trait substitution through nearby mutations is only successful when this substitution moves the population closer to the equilibrium. Mathematically, the adaptive dynamics of mutation and trait substitution is modeled here by the *canonical equation* (Marrow et al., 1996), a one-dimensional dynamical system (see Section 4) whose stable equilibria are characterized by the static convergence stable conditions of Christiansen (1991) (also called the *m*-stability concept in Taylor, 1989). Combined with a further condition (often called the ESS or uninvadability condition) that guards against the monomorphism being successfully invaded by an evolving dimorphism through a process now referred to as evolutionary branching (Doebeli and Dieckmann, 2000), we obtain the solution concept of *continuously stable strategy* (CSS) introduced by Eshel (1983).

From our perspective, adaptive dynamics and its canonical equation are approximate descriptions of how the mean of the distribution of individual behaviors evolves and do not adequately model the spread of the distribution. In this paper, we use the replicator equation with a continuous trait space to model the evolution of the probability distribution (i.e. probability measure) of individual behaviors. The CSS and/or convergence stability conditions are then heuristic tools that at best can suggest when the distribution will evolve to a monomorphism (i.e. to a Dirac delta distribution in measure theoretic terminology). In fact, Cressman and Hofbauer (2005) have shown the relevance of the CSS concept (and the closely related concept of a *neighborhood invader strategy* (NIS) of Apaloo, 1997) for stability of monomorphisms in the measure dynamics of a one-dimensional continuous trait space. Specifically, a non-CSS monomorphism is unstable in the measure dynamic and, conversely, a CSS is dynamically stable if the initial distribution of individual behaviors is close to the CSS and satisfies an additional technical requirement concerning the strategies present in the population (i.e. the support of this distribution).³ Of

²The matrix-ESS terminology will help avoid confusion with the term ESS as it has been used in the literature with continuous trait spaces.

³See Section 5 for further details of this technical requirement that successful mutant monomorphisms in the adaptive dynamics approach are available for trait substitution. Without this requirement, the stronger NIS conditions are needed to guarantee dynamic stability.

particular interest in proving these results is the technique of iterated elimination of strictly dominated strategies that is borrowed from classical game theory and also used in the stability analysis of evolutionary game theory applied to matrix games (e.g. Samuelson and Zhang, 1992) but seems not to have been used previously in the adaptive dynamics literature.

The main purpose of this paper is then to examine whether static extensions of the one-dimensional adaptive dynamics concepts continue their relevance for the measure dynamic model when the trait space is multi-dimensional. We begin in Section 2 by briefly developing the replicator equation on a continuous trait space and introducing essential notation used throughout the paper concerning this measure dynamics and the underlying fitness functions. Sections 3 and 4 provide valuable insight by fully analysing the replicator equation in the special case of quadratic fitness functions and normal distributions (Section 3) and then relating these results to potential static extensions of the CSS concept to multi-dimensional adaptive dynamics in Section 4 (see Theorems 5 and 6 there). Section 5 considers the general case of arbitrary fitness functions and probability distributions in the multi-dimensional setting. Unfortunately, our results here do not give as thorough a static characterization of stability for the replicator equation as that available through the one-dimensional analysis of Cressman and Hofbauer (2005). Although we are able to obtain necessary conditions related to adaptive dynamics for stability of monomorphisms in large classes of measure dynamic models as well as sufficient conditions for others, an exhaustive classification is beyond the reach of our current techniques. The final section discusses these shortcomings as it summarizes the positive aspects of our classification.

2. The replicator dynamics on the space of probability measures

The probability measure dynamics is the extension of the replicator equation originally defined for matrix games with a finite trait space (Taylor and Jonker, 1978). In general, individuals are assumed to play a strategy s in a fixed trait space S and the population state is given by a probability measure P on a measurable space (S, \mathcal{B}) . If $A \in \mathcal{B}$, $P(A)$ is interpreted as the proportion of individuals in the population who are using strategies in the set A .

For our model of a multi-dimensional continuous trait space, S will be a Borel subset of \mathbf{R}^n and \mathcal{B} will be the Borel subsets of S (i.e. the σ -algebra of the Borel sets of \mathbf{R}^n intersected with S and so P is a Borel measure).⁴ Let $\mathcal{A}(S)$ denote the set of probability measures with respect to (S, \mathcal{B}) . Since P is a Borel measure, there is a unique

⁴In fact, S will typically have further topological properties such as being convex and open (or closed with nonempty interior). The Borel subsets of \mathbf{R}^n form the smallest σ -algebra containing the open subsets of \mathbf{R}^n .

(relatively) closed subset of S , called the *support* of P , such that the measure of its complement is 0 but every open set that intersects it has positive measure (Royden, 1988).

The measure dynamics on $\Delta(S)$ (see Eq. (1) below) is given in terms of the fitness (or expected payoff) $\pi(s, P) = \pi(\delta_s, P)$ of an individual using strategy s when the population is in state P . Here, for a given $s \in S$, δ_s denotes the Dirac delta measure that assigns unit mass to $\{s\}$. We will assume throughout that the fitness $\pi(s, P)$ is given through a continuous real-valued payoff function $\pi : S \times S \rightarrow \mathbf{R}$ by $\pi(s, P) = \int_S \pi(s, y)P(dy)$. In particular, matrix games that assume random pairwise interactions and a finite trait space S may be put in this form.⁵ The mean payoff to a random individual in the population with state P is then $\pi(P, P) \equiv \int_S \pi(\delta_s, P)P(ds)$.

We assume the *replicator equation* Eq. (1) describes how the population state evolves (i.e. its solutions define trajectories P_t in $\Delta(S)$).

$$\frac{dP}{dt}(A) = \int_A (\pi(\delta_s, P) - \pi(P, P))P(ds). \quad (1)$$

Heuristically, this dynamic increases the probability of those sets of strategies A that have a higher expected payoff than the mean payoff to a random individual in the population. It has been shown (Oechssler and Riedel, 2001) that there is a unique solution that satisfies this dynamics for all positive t given any initial probability measure P_0 with compact support when $\pi(s, y)$ is continuous.⁶ Here A is a Borel subset of S and dP/dt at time t is defined to be $\lim_{h \rightarrow 0} (P_{t+h} - P_t)/h$ with respect to the variational norm (i.e. $\lim_{h \rightarrow 0} \|\frac{dP}{dt}(A) - (P_{t+h} - P_t)/h(A)\| = 0$ where $\|\cdot\|$ is the variational norm as in Oechssler and Riedel, 2001). Furthermore, the support of P_t is the same as for P_0 for all $t \geq 0$. A population state P^* is an equilibrium of Eq. (1) if and only if $\pi(\delta_s, P^*) - \pi(P^*, P^*) = 0$ for all $s \in \text{supp}(P^*)$.

Our primary aim in this paper is the investigation of the convergence and stability properties (especially related to monomorphic equilibrium population states δ_s) of the measure dynamics (1). Heuristically, dynamic stability of P^* refers to the question whether P_t stays close and/or evolves to P^* if the initial P_0 is chosen appropriately in $\Delta(S)$. From Oechssler and Riedel (2002), it is clear that the answers to the stability question depend critically on the concept of closeness of probability measures (i.e. on the topology used for the space of Borel probability measures), when the trait space is not a finite subset of \mathbf{R}^n .

We feel the *weak topology* captures best the essence of evolutionary convergence in our biological systems. This topology will mostly be applied to neighborhoods of monomorphic P^* . In general, for a probability measure P^*

with finite support $\{x_1, \dots, x_m\}$, we can take ε -neighborhoods in the weak topology to be of the form

$$\{Q \in \Delta(S) : |Q(B_\varepsilon(x_i)) - P^*({x_i})| < \varepsilon \forall i = 1, \dots, m\},$$

where $B_\varepsilon(x)$ is the open ball of radius ε centered at x . In particular, two monomorphisms δ_{x_1} and δ_{x_2} are within ε of each other if and only if the Euclidean distance between these points is less than ε . In the following all topological notions in $\Delta(S)$ are taken for this weak topology, unless otherwise stated.

2.1. The fitness function $\pi(s, y)$

For the multi-dimensional continuous trait space, we assume S is the closure of an open connected subset of \mathbf{R}^n that contains the origin in its interior. In fact, we often assume S is star-shaped with respect to the origin (i.e. if $x \in S$, then so does the line segment joining 0 to x). We are particularly interested in the stability of the monomorphism δ_0 . To this end, consider the Taylor expansion of $\pi(x, z)$ about $(0, 0) \in \mathbf{R}^{2n}$

$$\begin{aligned} \pi(x, z) = & \pi(0, 0) + \nabla_1 \pi \cdot x + \nabla_2 \pi \cdot z + \frac{1}{2}[x \cdot (\nabla_{11}^2 \pi)x \\ & + 2x \cdot (\nabla_{12}^2 \pi)z + z \cdot (\nabla_{22}^2 \pi)z] + h.o.t., \end{aligned}$$

where for $i \in \{1, 2\}$, $\nabla_i \pi$ is the gradient vector of π at the origin in the i th variable (e.g. $(\nabla_1 \pi)_k = (\partial \pi(x, z) / \partial x_k)|_{(x,z)=(0,0)}$) and $\nabla_{ij}^2 \pi$ is the $n \times n$ matrix with entries the appropriate second-order partials.

Each monomorphism is a rest point of Eq. (1). Their stability in the weak topology requires the monomorphism be a NE of the payoff function restricted to the game with nearby strategies (e.g. 0 is such a NE if $\pi(x, 0) \leq \pi(0, 0)$ for all x sufficiently close to 0). Since 0 is an interior point, this implies the gradient $\nabla_1 \pi$ is the zero vector and $x \cdot (\nabla_{11}^2 \pi)x \leq 0$ for all $x \in \mathbf{R}^n$. In fact, we will assume the symmetric Hessian matrix $\nabla_{11}^2 \pi$ is negative definite throughout to avoid technical issues. That is, we assume 0 is a strict NE of the restricted game.⁷ Since $\nabla_{11}^2 \pi$ is symmetric, we can diagonalize it by an orthogonal transformation and then all diagonal entries $-d_k$ are negative. Furthermore, a change of variables (that replaces x_k with $\sqrt{d_k}x_k$ and takes the payoffs with respect to these new variables) allows us to assume $\nabla_{11}^2 \pi = -2I$ where I is the $n \times n$ identity matrix. Without loss of generality, the fitness function can then be written in these new coordinates as

$$\begin{aligned} \pi(x, z) = & \pi(0, 0) + \nabla_2 \pi \cdot z - x \cdot x + x \cdot Bz \\ & + \frac{1}{2}z \cdot (\nabla_{22}^2 \pi)z + h.o.t. \end{aligned} \quad (2)$$

(i.e. $\nabla_{11}^2 \pi = -2I$ and $\nabla_{12}^2 \pi = B$).

⁵In fact, Bomze and Pötscher (1989) argue that the existence of such a $\pi(s, y)$ for an arbitrary trait space S means the evolutionary game can be interpreted as being based on pairwise interactions. It is only the form of $\pi(s, P)$ that is important to us, not whether players are competing pairwise.

⁶If $\pi(s, y)$ is not continuous or P_0 does not have compact support, one must be careful that the desired integrals are defined. The latter concern is discussed further for the normal distributions of Section 3.

⁷The condition $\pi(x, 0) < \pi(0, 0)$ is related to the concepts of uninvadability and evolutionarily stable strategy (ESS) as used in adaptive dynamics (Marrow et al., 1996; Vincent et al., 1996). We especially avoid this latter terminology since the ESS description is overused in the literature and may have different interpretations for different readers. On the other hand, strict NE seems to have a universally accepted meaning.

In terms of the Taylor expansion, the replicator equation (1) can then be rewritten as

$$\frac{dP}{dt}(A) = \int_S \int_S \int_A (\pi(x, z) - \pi(y, z)) P(dx) P(dy) P(dz), \quad (3)$$

where the integrand is given by

$$\pi(x, z) - \pi(y, z) = (y - x) \cdot [x + y - Bz + h.o.t.]. \quad (4)$$

In particular, the constant and linear terms in the fitness function are irrelevant for the dynamic analysis.

3. The replicator equation with normal distributions and quadratic fitness functions

In this section, we analyse the replicator equation when the higher-order terms are ignored in Eq. (4) and the initial probability measure P_0 is the (multivariate) normal distribution $N(m, C)$ with mean vector $m \in \mathbf{R}^n$ and covariance matrix $C \in \mathbf{R}^{n \times n}$. From Section 2.1, we may assume fitness has the form of the quadratic function

$$\pi(x, z) = -x \cdot x + x \cdot Bz \quad (5)$$

for $x, z \in S = \mathbf{R}^n$ and B an $n \times n$ matrix.

We proceed as follows. The first step is to show that the class of normal distributions is forward invariant under the replicator equation. Therefore, the infinite-dimensional measure dynamics is reduced to a finite-dimensional system of $n + n^2$ ordinary differential equations for the mean and covariance matrix. These facts are stated in Theorem 1 below where it is also apparent that the dynamics of the covariance matrix does not depend on the mean vector. The next step is to obtain the explicit solution (Theorem 2) for the evolution of the covariance matrix. Substitution of this solution into the dynamics for the mean results in a system of linear differential equations with time varying coefficients. The stability analysis of this system for the equilibrium δ_0 (i.e. for the limit of the normal distributions $N(0, C)$ as C approaches the zero matrix) is summarized in Theorems 3 and 4 in terms of the matrix B .

Theorem 1. *The class of normal distributions is forward invariant under the replicator dynamics (1). Assume that the initial distribution is normal, $P_0 = N(m, C)$ for a mean vector $m \in \mathbf{R}^n$ and a covariance matrix $C \in \mathbf{R}^{n \times n}$. Then the solution of the replicator dynamics starting at P_0 is given by $P_t = N(m(t), C(t))$, where the mean and the covariance matrix solve the initial value problem*

$$m'(t) = C(t)(B - 2I)m(t) \quad (6)$$

$$C'(t) = -2C(t)C(t). \quad (7)$$

with $m(0) = m$ and $C(0) = C$.

The proof of Theorem 1 (and most other theorems) is in the Appendix. Since the dynamical system (7) for the covariance matrix does not depend on the mean vector, we may study this system of differential equations on its own. The following theorem gives the explicit solution which can be easily verified.

Theorem 2. *For any initial positive semidefinite matrix $C(0)$, the solution of Eq. (7) is given by*

$$C(t) = C(0)(I + 2C(0)t)^{-1}. \quad (8)$$

Note that Eq. (8) is well defined for all $t \geq 0$. Alternatively, one can write the solution in the following way.

$$C(t) = O^T D(t) O,$$

where O is an orthogonal matrix such that $OC(0)O^T = D$ for some diagonal matrix D (and O^T denotes the transpose of O) and $D(t)$ is the diagonal matrix with entries

$$D_{ii}(t) = \frac{D_{ii}}{1 + 2tD_{ii}}. \quad (9)$$

In particular, the covariance matrix $C(t)$ converges to the zero matrix, and $C(t) = (1/2t)I + O(1/t^2)$ as $t \rightarrow \infty$ whenever the initial condition $C(0)$ is positive definite. By inserting solution (8) into Eq. (6), the mean evolves according to the time-dependent linear differential equation

$$\frac{dm}{dt} = C(0)(I + 2C(0)t)^{-1}(B - 2I)m(t). \quad (10)$$

After changing the time-scale ($t + 1 = e^{2\tau}$) this differential equation becomes asymptotically autonomous

$$\frac{dm}{d\tau} = (B - 2I)m(\tau) + R(\tau)m(\tau) \quad (11)$$

with exponentially decreasing remainder term $R(\tau) = (2C(0) - I)(I + 2(e^{2\tau} - 1)C(0))^{-1}(B - 2I)$. Therefore the eigenvalues of $B - 2I$ determine the asymptotic behavior of $m(t)$ and imply the following theorem (see its proof in the Appendix for further details).

Theorem 3. *Consider the replicator equation (1) restricted to the class of normal distributions with quadratic fitness functions as in Eq. (5).*

1. *If every eigenvalue of the matrix $B - 2I$ has negative real part, then δ_0 attracts all normal distributions P_0 of the form $N(m, C)$ with positive definite symmetric covariance matrix C (i.e. $C(t) \rightarrow O$ and $m(t) \rightarrow 0$ as $t \rightarrow +\infty$). Furthermore, δ_0 is asymptotically stable with respect to all such P_0 with initial covariance matrix C orthogonally similar to a positive diagonal matrix D (i.e. $OCO^T = D$) satisfying $\min\{D_{ii}/D_{jj} \mid 1 \leq i, j \leq n\} \geq \varepsilon$ for some $\varepsilon > 0$.*
2. *If some eigenvalue of $B - 2I$ has positive real part, then δ_0 is unstable: for all positive definite C there are m arbitrarily close to 0 such that along the solution with initial $P_0 = N(m, C)$ one has $\|m(t)\| \rightarrow \infty$ as $t \rightarrow \infty$.⁸*

It is important to note that, in part 1 of Theorem 3, we do not state that δ_0 is asymptotically stable with respect to all P_0 of the form $N(m, C)$ with positive definite symmetric

⁸ $\|m\|$ is the Euclidean norm of $m \in \mathbf{R}^n$.

covariance matrix C . In fact, this asymptotic stability requires stronger assumptions on B .

To see this, we consider degenerate normal distributions $N(m, C)$ with a positive semidefinite covariance matrix C with m in the range of C (i.e. Gaussian distributions confined to a linear subspace of \mathbf{R}^n). Then Eq. (11) is replaced by

$$\frac{dm}{d\tau} = P(B - 2I)m(\tau) + PR(\tau)m(\tau), \quad (12)$$

where $P = \lim_{t \rightarrow \infty} 2tC(0)(I + 2tC(0))^{-1}$.⁹ Then P is a projection matrix (i.e. $P = P^T = P^2$) since its eigenvalues are 0 and 1 only.

Note that two (possibly degenerate) normal distributions are close in the weak topology if and only if their means and covariances are close in the Euclidean metric. Since there are normal distributions with a positive definite covariance matrix arbitrarily close to these degenerate ones, stability requires every eigenvalue of $P(B - 2I)$ to have negative real part. This condition implies the following result on instability.

Theorem 4. *If $B - 2I$ is not negative semidefinite (i.e. $x \cdot Bx > 2\|x\|^2$ for at least one $x \in \mathbf{R}^n$) then δ_0 is unstable for Eq. (1) restricted to normal distributions.*

It is instructive to compare Theorems 3 and 4 in the case where every eigenvalue of the matrix $B - 2I$ has negative real part but $B - 2I$ is not negative semidefinite (in particular, $B - 2I$ cannot be a symmetric matrix). In the proof of Theorem 4 we construct rank one matrices $C(0)$ and means $m(0)$ such that $m(t) \rightarrow \infty$ as $t \rightarrow +\infty$, i.e. there is divergence for normal distributions concentrated on certain lines through 0. By continuous dependence on initial conditions in Eqs. (6) and (7), there are positive definite covariance matrices $\tilde{C}(0)$ and means $\tilde{m}(0)$ close by such that $\|\tilde{m}(t)\|/\|\tilde{m}(0)\|$ becomes arbitrarily large for some t . Hence δ_0 is unstable even for normal distributions with positive definite covariance matrix. However, in this case, $\tilde{m}(t)$ will eventually converge to 0 after this long excursion away by Part 1 of Theorem 3.1.

This phenomenon—attractivity without stability—requires a multi-dimensional trait space.

4. Multi-dimensional adaptive dynamics, the canonical equation and CSS

As stated in the Introduction, a main purpose of this paper is to examine the relevance (for the dynamic stability of the replicator equation with multi-dimensional continuous strategy space) of static extensions of the one-dimensional stability conditions developed by adaptive dynamics (e.g. the CSS and NIS concepts). The compar-

isons developed in this section continue to be based on a quadratic payoff function $\pi(x, y)$.

The *canonical equation* (13) from adaptive dynamics for the evolution of a (monomorphic) population with mean m through mutation and trait substitution was developed by Dieckmann and Law (1996). Following Meszena et al. (2001) (see also Leimar, 2005), this takes the form (in our notation)

$$m'(t) = \frac{1}{2}u(m)N(m)C(m)\nabla_1\pi(x, m)|_{x=m}. \quad (13)$$

Here $u(m)$ and $N(m)$ are positive real-valued functions giving the mutation rate and the equilibrium population size respectively at mean m . These can be ignored in analysing the limiting behavior of the canonical equation since they do not affect the evolutionary trajectory but only the speed of evolution along this trajectory (and so $\frac{1}{2}u(m)N(m)$ will be deleted from now on). More importantly, the covariance matrix $C(m)$ (which now characterizes the expected mutational effects in different directions from m and does affect its evolution) is assumed to depend only on m . In the earlier formulation of adaptive dynamics by Hofbauer and Sigmund (1990) the symmetric and positive definite matrix $C(m)$ comes from a Riemannian metric on the trait space.

To compare Eq. (13) to the replicator dynamics, assume $m = 0$ is a strict NE in the interior of the trait space as in Section 2.1. Then $\nabla_1\pi(x, 0)|_{x=0} = 0$ and so $m = 0$ is an interior equilibrium of Eq. (13). With the same change of variables that led to Eq. (2) in Section 2.1, the canonical equation becomes

$$m'(t) = C(m(t))(B - 2I)m(t). \quad (14)$$

$m = 0$ is called *convergence stable* (with respect to $C(m)$) if it is asymptotically stable under Eq. (14).

A quick glance at Eq. (6) shows the canonical equation is quite closely related to the evolution of the mean for normal distributions under Eq. (1) with quadratic fitness functions. When $C(m(t)) = c(t)I$ for some positive function $c(t) > 0$, the two dynamical systems have the same trajectories for the mean although the mean evolves much slower under the replicator equation through the change in time-scale (given by $t + 1 = e^{2\tau}$) as the covariance approaches the zero matrix. In general, the only difference mathematically is that Eq. (14) is an autonomous system of differential equations whereas Eq. (6) is not. As we will see, this difference has important consequences for multi-dimensional trait space on how convergence stability is related to dynamic stability of Eq. (6) where the covariance matrix $C(t)$ is given explicitly in Theorem 2.

For a one-dimensional trait space, convergence stability is independent of the choice of $C(m)$. That is, $m = 0$ is asymptotically stable with respect to the canonical Eq. (14) for one choice of positive variance as a function of m if and only if it is for any other choice. In fact, a one-dimensional strict NE that is convergence stable is called a *Continuously Stable Strategy* (CSS), a concept introduced by Eshel (1983). Furthermore, $m = 0$ is a CSS if and only if δ_0 is

⁹Inserting $C(0) = PC(0)$ in Eq. (10) leads to Eq. (12) instead of Eq. (11) with $PR(\tau)$ exponentially decreasing when restricted to m in the range of C .

asymptotically stable under (6).¹⁰ Unfortunately, this correspondence does not extend beyond one dimension as illustrated by the following example.

Example 1. Consider the two-dimensional trait space \mathbf{R}^2 with quadratic fitness function (5) and

$$B = \begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix},$$

where b, c are both positive. By Theorem 3, δ_0 is attractive under Eq. (1) for the class of normal distributions if and only if the eigenvalues of $B - 2I$, $\lambda_{1,2} = -2 \pm \sqrt{bc}$, are both negative, i.e., $bc < 4$. Also, by Theorem 4, δ_0 is unstable for this class if $B - 2I$ is not negative semidefinite, i.e., $|b + c| < 4$.

Example 1a. Suppose $C(m)$ is the constant symmetric matrix

$$C = \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix}$$

for all m where α is a fixed parameter satisfying $|\alpha| < 1$ so that C is positive definite. Then $m = 0$ is globally asymptotically stable under Eq. (14) if and only if every eigenvalue of $C(B - 2I)$ has negative real part.

With $b = 1/4$ and $c = 4$, the eigenvalues of $B - 2I$ are $-1, -3$ and so δ_0 is attractive for the replicator dynamics restricted to the normal distributions if $|\alpha| < 1$. On the other hand, the sum of the eigenvalues of $C(B - 2I)$ is the trace $-4 + 17\alpha/4$ of

$$\begin{pmatrix} -2 + 4\alpha & 1/4 - \alpha \\ -2\alpha + 4 & \alpha/4 - 2 \end{pmatrix}.$$

Thus, some eigenvalue has positive real part if $16/17 < \alpha < 1$ and so $m = 0$ is not attractive for the canonical equation (14) with constant covariance matrix parameterized by these α .

The mathematical reason for this difference between properties of the replicator equation and the canonical equation (see Theorem 6 below) is that $B - 2I$ is not negative definite for $b = 1/4$ and $c = 4$ since $|b + c| > 4$. Furthermore, even if $B - 2I$ were negative definite, the asymmetry of B implies there is a continuous choice $C(m)$ depending on m for which $m = 0$ is unstable.

Example 1b. For an explicit example of this latter phenomenon, take $b = 1$ and $c = 1/2$ (so $B - 2I$ is negative definite) with

$$C(m) = \begin{pmatrix} \alpha^2 & \alpha\beta \\ \alpha\beta & \beta^2 \end{pmatrix},$$

where

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \equiv \begin{pmatrix} -14 & 42 \\ -39 & 16 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}.$$

Substitution into Eq. (14) yields

$$\begin{aligned} \begin{pmatrix} m'_1 \\ m'_2 \end{pmatrix} &= \begin{pmatrix} \alpha^2 & \alpha\beta \\ \alpha\beta & \beta^2 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 1/2 & -2 \end{pmatrix} \begin{pmatrix} -14 & 42 \\ -39 & 16 \end{pmatrix}^{-1} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ &= \frac{1}{202} \begin{pmatrix} \alpha^2 & \alpha\beta \\ \alpha\beta & \beta^2 \end{pmatrix} \begin{pmatrix} 1 & 10 \\ -10 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ &= \frac{\alpha^2 + \beta^2}{202} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \end{aligned}$$

Thus,

$$\begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} = \frac{\alpha^2 + \beta^2}{202} \begin{pmatrix} -14 & 42 \\ -39 & 16 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

and so $m = 0$ is unstable for Eq. (14) since the trace of this last 2×2 matrix is positive.¹¹ On the other hand, δ_0 is globally asymptotically stable for the replicator dynamics restricted to the normal distributions in the sense of Theorem 3, Part 1.

The two choices of explicit parameters in Examples 1a and 1b above illustrate two methods to extend the one-dimensional CSS conditions to multi-dimensions. The more common approach (Meszena et al., 2001) is to consider $m^* \in \mathbf{R}^n$ a multi-dimensional CSS if it is a strict NE that is convergence stable with respect to any choice of constant positive definite symmetric covariance matrix C . Translating m^* to the origin, this is equivalent to asserting $C(B - 2I)$ has every eigenvalue with negative real part for all choices of C . Hines (1980b); Cressman and Hines (1984) (see also Leimar, 2005) show this is true if and only if $B - 2I$ is negative definite, which in the above example is the condition $|b + c| < 4$. The negative definiteness of $B - 2I$ is also equivalent to (multi-dimensional) m -stability introduced by Lessard (1990).

In another approach, Leimar (2005) considers a more restrictive notion of CSS by allowing all continuously varying $C(m)$ that are positive definite but not necessarily symmetric, a condition Leimar called *absolute convergence stability* (see also the concept of a Darwinian demon in Leimar, 2001). He then showed this condition is equivalent to $B - 2I$ being symmetric and negative definite, a similar result as that illustrated in Example 1b. The essential properties for this specific example are that $A \equiv \begin{pmatrix} -14 & 42 \\ -39 & 16 \end{pmatrix}$ has an eigenvalue of positive real part and that $A^T(B - 2I)$ is positive definite.

¹⁰The condition for asymptotic stability in both dynamics is that the only entry $b - 2$ of the matrix $B - 2I$ is negative (i.e. $b < 2$). Here we ignore the degenerate case with $b = 2$. Similarly, the above definition of CSS ignores the possibility that a non-strict NE can be a CSS, a situation we also view as degenerate.

¹¹ $C(m)$ is only positive semidefinite. Positive terms can be added to the diagonal of $C(m)$ to make it positive definite, at least for $m \neq 0$. If these terms are of the form $\alpha(\alpha^2 + \beta^2)$ (i.e. they go to zero faster than $\alpha^2 + \beta^2$), they will not affect the instability of $m = 0$ under Eq. (14).

By the following definition, we follow the first perspective of Lessard (1990); Meszena et al. (2001) in this paper.

Definition 1. $m^* \in \mathbf{R}^n$ is a multi-dimensional CSS if it is a strict NE and convergence stable with respect to any choice of constant positive definite symmetric covariance matrix C .

That is, we define a strict NE to be a CSS if $B - 2I$ is negative definite, i.e., $x \cdot Bx < 2\|x\|^2$ for all $x \neq 0$. There are several reasons for this choice. Not only is mutational covariance near a monomorphic equilibrium assumed to be constant in most treatments of adaptive dynamics (Vincent et al., 1993; Metz et al., 1996; Meszena et al., 2001), it is also a standard assumption in earlier treatments of evolution of strategy distributions from game-theoretic models where payoff functions are often assumed bilinear as in $\pi(x, y) = x \cdot By$ (Hines, 1980b).¹² For us, another important reason is that this definition of CSS is the most relevant condition for dynamic stability of the general replicator equation (1) analysed in the following section (see Theorem 12 there).

The above example and/or the proof of Theorem 3 show that dynamic stability depends critically on the evolution of the covariance matrix $C(m(t))$ and/or $C(t)$. However, if $B - 2I$ is symmetric, all concepts are equivalent. That is, we have the following.

Theorem 5 (Symmetric B). Suppose B is symmetric. The following four statements are equivalent.¹³

- (1) δ_0 is attractive under Eq. (1) for the class of non-degenerate normal distributions.
- (2) δ_0 is Lyapunov stable under Eq. (1) for the class of all normal distributions.
- (3) $m = 0$ is a CSS.
- (4) $m = 0$ is a strict NE and absolutely convergence stable.

The reason for this equivalence is that, for symmetric matrices, negative definiteness is equivalent to all eigenvalues being negative. The stability assertion in statement 2 follows from the *Fundamental Theorem of Natural Selection* (see Oechssler and Riedel, 2002, Theorem 1, or Cressman and Hofbauer, 2005, Section 4.2). Specifically, mean fitness $\tilde{\pi}(P, P)$ is a Lyapunov function for Eq. (1) for the equivalent payoff function $\tilde{\pi}(x, y) = -x \cdot x + x \cdot By - y \cdot y$ which is symmetric (i.e. $\tilde{\pi}(x, y) = \tilde{\pi}(y, x)$) and $P = \delta_0$ is a maximizer of $\tilde{\pi}(P, P)$ if and only if $B - 2I$ is negative definite.

¹²See however Hines (1980a) where it is shown that non-constant variances arise quite naturally and can play an important role in the stability analysis. Effects of evolving (co)variances are important in models of quantitative genetics as well (e.g. Burger, 2000) where the mean strategy dynamics is similar to the canonical equation, although here variances are again sometimes assumed constant (Abrams, 2001).

¹³These equivalences ignore degenerate possibilities. For example, we ignore situations where relevant eigenvalues have zero real part.

On the other hand, if B is not symmetric, none of the statements are equivalent by Examples 1a and 1b. We then have the following theorem.

Theorem 6. Suppose B is not symmetric and $m = 0$ is a strict NE. Then

- (1) δ_0 is attractive and asymptotically stable (as in Theorem 3, Part 1) under Eq. (1) for the class of normal distributions if and only if every eigenvalue of $B - 2I$ has negative real part.
- (2) $m = 0$ is a CSS if and only if $B - 2I$ is negative definite.
- (3) $m = 0$ is not absolutely convergence stable.

Remark. By Theorem 6 and the discussion following Theorem 4, the multi-dimensional CSS concept emerges by requiring asymptotic stability with respect to all positive semidefinite covariance matrices with one-dimensional range (i.e. a line through the origin). By Theorems 3 and 4 restricted to each such line, we have an independent proof of the result (Lessard, 1990; Meszena et al., 2001) that the multi-dimensional CSS concept is equivalent to the one-dimensional CSS conditions for each line through the origin.

5. Stability of monomorphisms under the replicator equation

The explicit analysis of the replicator dynamics in Section 3 relies heavily on the assumptions the payoff function $\pi(x, y)$ is quadratic and the initial population is normally distributed with mean m . Such normal distributions are one way to model aggregate individual mutations for a monomorphic population at m . An underlying assumption of the replicator equation (1) is that each individual reproduces offspring with its same trait and at a rate equal to its fitness. If this reproduction of clones is subject to small shocks from m that are independent of each other, the Central Limit Theorem can be used to conclude the initial traits will be normally distributed after such a shock. So here mutation is treated indirectly by considering variations of initial conditions.

In the remainder of the paper, we consider other initial distributions (that can also be given a mutational interpretation) and arbitrary payoff functions. We will be most interested in convergence and stability properties of monomorphisms δ_{m^*} under Eq. (1) for initial distributions whose support is close to m^* to reflect the adaptive dynamics assumption that trait substitution involves only nearby mutations. This means we cannot simply use the weak topology on $\Delta(\mathbf{R}^n)$ (or on $\Delta(S)$ for that matter) since P may be close to δ_{m^*} in the weak topology and still have support all of \mathbf{R}^n (e.g. a normal distribution with mean 0 and small variance is close to δ_0 in the weak topology). That is, for stability with respect to the weak topology, one must also account for possibly large payoff effects of (admittedly rare) interactions involving an individual whose strategy is far from that of the monomorphic

population. To avoid this problem, we concentrate instead on convergence to δ_{m^*} in the weak topology for initial distributions whose support is close to m^* (see the concept of neighborhood attractivity in the following definition). It is important to note that neighborhood attractivity also requires m^* to be in the support of these initial distributions since the replicator dynamics (1) in a continuous trait space shares the same property of its analog for a finite number of traits; namely, $\text{supp } P_t = \text{supp } P_0$ for all $t \geq 0$ (Bomze, 1991). The main advantage of this concept is that the Taylor expansion of the payoff function about the monomorphism can be used (specifically, up to quadratic terms) as a good approximation.

Remark. Alternatively, we could modify the weak topology by defining P close to P^* if P is close to P^* in the weak topology and the support of P is close to that of P^* . In fact, when the trait space is one dimensional, this is the approach taken by Eshel and Sansone (2003) who called this the *maximal shift topology*. However, every monomorphism $P^* = \delta_{m^*}$ is then automatically stable in this modified weak topology since P is close to δ_{m^*} if and only if its support is close to m^* .¹⁴ However, P_t cannot converge to δ_{m^*} in this new topology since the support of P_t does not converge to $\{m^*\}$. That is, if δ_{m^*} is neighborhood attractive, then it immediately enjoys stability properties under this new topology but convergence must still be taken with respect to the weak topology. For this reason, we prefer not to formally define a modified weak topology.

Definition 2. Let Q be a set of probability distributions, whose support contains that of P^* , that is forward invariant under Eq. (1). P^* is *attractive for* Q if P_t converges to P^* in the weak topology for every $P_0 \in Q$. $P^* = \delta_{m^*}$ is *neighborhood attractive for* Q if, for some neighborhood U of m^* in \mathbf{R}^n , P_t converges to P^* in the weak topology for every $P_0 \in Q$ with $\text{supp } P_0 \subset U$.

Clearly, neighborhood attractivity depends critically on the choice of Q . For instance, in the trivial case that δ_{m^*} is the only distribution in Q with support close to $\{m^*\}$, δ_{m^*} is neighborhood attractive by default. Sections 5.1 and 5.2 examine the attractivity properties of Definition 2 for two choices of Q that are more important. In either case, it is assumed that Q contains measures whose support is a small closed neighborhood of $m^* = 0$ in the Euclidean topology. As shown in Lemma A.1 in the Appendix, the attractivity concept in Definition 2 then requires at a minimum that m^* be a NE locally (i.e. $\pi(m^*, m^*) \geq \pi(x, m^*)$ for all x near m^* , see also Alós-Ferrer and Ania, 2001). To avoid technical complications, we again assume throughout Section 5 that m is in fact a local strict NE as determined by the second-order Taylor expansion of π . Furthermore, we assume that the trait space has been parameterized in such a way that $m^* = 0$ is in the interior of S and the Taylor expansion of

$\pi(x, y)$ up to quadratic terms is given by Eq. (5). To repeat, we make the following assumption.

Assumption. $m^* = 0$ is a local strict NE (i.e. $\pi(x, 0) < \pi(0, 0)$ for all x sufficiently close (but not equal) to 0) and the Taylor expansion of $\pi(x, y)$ about $(0, 0)$ is $\pi(x, y) = -x \cdot x + x \cdot B y$ up to second-order terms.

5.1. Dynamic stability, neighborhood superiority and NIS

In this section, we seek the strongest type of attractivity possible by taking $Q = Q_0 := \{P \in \Delta(S) : 0 \in \text{supp } P\}$. Consider the trajectory in $\Delta(S)$ for initial P_0 with support $\{0, x\}$ and $x \in S$. The replicator equation (1) for such P_0 is

$$\begin{aligned} \frac{dP}{dt}(\{0\}) &= P(\{0\})(\pi(\delta_0, P) - \pi(P, P)) \\ &= P(\{0\})P(\{x\})[(\pi(0, 0) - \pi(x, 0))P(\{0\}) \\ &\quad + (\pi(0, x) - \pi(x, x))P(\{x\})] \\ &= P(\{0\})P(\{x\})[x \cdot x P(\{0\}) \\ &\quad + (x \cdot x - x \cdot B x)P(\{x\})] + h.o.t. \end{aligned} \tag{15}$$

If $x \cdot x - x \cdot B x < 0$ for some $x \neq 0$, then $dP/dt(\{0\}) < 0$ if x is sufficiently close to 0 and $P(\{x\})$ is sufficiently close to 1. Since 0 is in the interior of S , we may assume $x \in S$. Thus, P_t does not converge to δ_0 and so δ_0 is not attractive for such dimorphic P_0 .

That is, neighborhood attractivity of δ_0 for Q_0 implies $B - I$ is negative definite.¹⁵ This negative definiteness condition is similar to the CSS condition of adaptive dynamics (see Definition 1 of Section 4). In fact, it is more closely related to the static condition called a *neighborhood invader strategy* (NIS) (McKelvey and Apaloo, 1995; Apaloo, 1997).¹⁶ The trait 0 is NIS if it has higher payoff against all nearby monomorphic populations δ_x than the expected payoff of x . That is, we have the following definition.

Definition 3. $m^* \in \mathbf{R}^n$ is a NIS if

$$\pi(m^*, x) > \pi(x, x) \tag{16}$$

for all other $x \in \mathbf{R}^n$ in a neighborhood of m^* .

From the Taylor expansion of a general payoff function $\pi(x, y)$ about $(0, 0)$, 0 is NIS if

$$\begin{aligned} \nabla_2 \pi \cdot x + \frac{1}{2} x \cdot (\nabla_{22}^2 \pi) x > \nabla_1 \pi \cdot x + \nabla_2 \pi \cdot x + \frac{1}{2} [x \cdot (\nabla_{11}^2 \pi) x \\ + 2x \cdot (\nabla_{12}^2 \pi) x + x \cdot (\nabla_{22}^2 \pi) x] + h.o.t. \end{aligned}$$

Since 0 is in the interior of S , $\nabla_1 \pi \cdot x = 0$ and so the NIS condition is that

$$x \cdot (\nabla_{11}^2 \pi) x + 2x \cdot (\nabla_{12}^2 \pi) x < 0.$$

¹⁵Throughout Section 5, we again ignore degenerate possibilities. Thus, here we assume $B - I$ is not negative semidefinite. There is a partial converse as well; namely, if $B - I$ is negative definite (i.e. $x \cdot x - x \cdot B x > 0$ for all x), then δ_0 attracts all dimorphic P with support containing 0.

¹⁶An NIS is also known as a *good invader* (Kisdi and Meszéna, 1995) and as satisfying (multi-dimensional) m^* -stability (Lessard, 1990). For a one-dimensional trait space, Eshel and Sansone (2003) proved the NIS condition is necessary for neighbourhood attractivity of δ_0 .

¹⁴This statement is no longer true for stability of a non-monomorphic distribution P^* (for dimorphisms, see Cressman, 2005).

By reparameterizing our trait space so that π is given by Eq. (2) (i.e. $\nabla_{11}^2 = -2I$ and $\nabla_{12}^2 = B$), the local strict NE 0 is NIS if and only if $B - I$ is negative definite. Furthermore, from Eq. (16) combined with the fact that $\pi(0, 0) > \pi(x, 0)$ for all other $x \in \mathbf{R}^n$ in a neighborhood of 0, $B - I$ is negative definite if and only if 0 strictly dominates all other nearby x in the two-strategy game based on the trait space $\{0, x\}$. This game-theoretic characterization that

$$\pi(0, z) > \pi(x, z)$$

for all $z \in \{0, x\}$ is important for comparison to the analogous characterization of the CSS condition in Section 5.2 (see Eq. (17) there).

Another game-theoretic characterization with attractivity consequences is given in terms of the following definition introduced by Cressman (2005) for continuous strategy spaces. Neighborhood superiority is closely connected to the concept of *evolutionarily robust* (Oechssler and Riedel, 2002) (also called *locally superior with respect to the weak topology* by Cressman and Hofbauer, 2005) whereby $\pi(P^*, P) > \pi(P, P)$ for all P sufficiently close to P^* in the weak topology.

Definition 4. The monomorphism $P^* = \delta_{m^*}$ is *neighborhood superior* if, for all other P with support sufficiently close to m^* , $\pi(P^*, P) > \pi(P, P)$.¹⁷

The following result summarizes the above discussion.

Theorem 7. *The following four statements are equivalent under our above Assumption for Section 5.*

- (1) δ_0 is neighborhood superior.
- (2) 0 is an NIS.
- (3) 0 strictly dominates all other nearby strategies x in the two-strategy game based on the trait space $\{0, x\}$.
- (4) $B - I$ is negative definite.

The only non-obvious implication in the above Theorem is that the first statement is implied by any one of the other three statements. This proof is Theorem 1 in Cressman (2005). An NIS need not be neighborhood superior if quadratic terms in the Taylor expansion do not determine the NIS conditions. Oechssler and Riedel (2002) provide the counterexample $\pi(x, y) = (x - y)^4 - 2x^4$ with a one-dimensional trait space (see also Eshel and Sansone, 2003).

Intuitively, a neighborhood superior P^* should be neighborhood attracting since P^* has a higher than average payoff at every nearby population state P .¹⁸ Unfortunately, we are only able to prove the following partial result for general payoff functions.

Theorem 8. *If δ_0 is neighborhood superior and P_0 has compact support sufficiently close to 0 and containing 0, then δ_0 is an ω -limit point of P_t .¹⁹*

In the special case of symmetric payoff functions (i.e. $\pi(x, y) = \pi(y, x)$), we have the following characterization, similar to Theorem 2 of Oechssler and Riedel (2002) and Theorem 4 in Cressman and Hofbauer (2005).

Theorem 9. *Suppose $\pi(x, y)$ is a symmetric payoff function (in particular, B is symmetric).²⁰ Then δ_0 is neighborhood attractive for the set Q_0 if and only if δ_0 is neighborhood superior.*

Remark. Attempts to extend Theorem 9 to general $\pi(x, y)$ have an interesting history. Oechssler and Riedel (2002) conjecture that Theorem 9 remains true for the weak topology when π is not symmetric (see their concept of evolutionarily robust). Eshel and Sansone, 2003 provide a proof of Theorem 9 for general $\pi(x, y)$ if the trait space is one dimensional (although we have been unable to follow all the details of this proof). If Q is taken as the set of all measures with $P(\{0\}) > 0$, Theorem 9 was proven by Bomze (1990) for $\pi(x, y) = \phi(x) + \phi(y)$, by Oechssler and Riedel (2002, Theorem 3) (see also Cressman, 2005, Theorem 1) for general $\pi(x, y)$.

5.2. Dynamic stability, dominance solvability and CSS

Section 5.1 illustrates the importance of the static NIS concept for stability of the replicator dynamics (1). We now turn to the relevance of the CSS concept. From the adaptive dynamics perspective, dynamics (15) models trait substitution from x to 0 in one step, whereas the canonical equation is built on the premise that mutation and trait substitution is a gradual process whereby x evolves to 0 in a sequence of many steps.

For a one-dimensional trait space, the canonical equation requires all traits between 0 and x be available for substitution and so we now assume the support of P_0 contains this interval. The heuristic condition (Eshel, 1983) for a strict NE at $m^* = 0$ to be a CSS amounts to replacing inequality (16) with

$$\pi(y, x) > \pi(x, x) \tag{17}$$

whenever y is close to x and between 0 and x . The Taylor expansion of $\pi(x, y)$ about $(0, 0)$ now yields a local strict NE satisfies Eq. (17) if and only if $b < 2$ where b is the only entry of the 1×1 matrix B in Eq. (5).

Cressman and Hofbauer (2005) were able to use an iterated strategy domination argument to show that, for any $|b| < 2$, δ_0 is attractive for initial distributions P_0 whose support is a (sufficiently small) interval S containing 0.

¹⁹ δ_0 is attractive for Q_0 if and only if δ_0 is the unique ω -limit point of P_t for all such $P_0 \in Q_0$.

²⁰We may assume $\pi(x, y)$ has the form $\pi(x, y) = -x \cdot x + x \cdot B y - y \cdot y + h.o.t.$ since the terms $\pi(0, 0) + \nabla_2 \pi \cdot y$ and $-y \cdot y$ (cf. Eq. (2)) are irrelevant for the replicator equation.

¹⁷Cressman (2005) analysed this concept for any P^* with finite support.

¹⁸For a finite trait space S , this intuition is equivalent to the notion of an evolutionarily stable strategy (i.e. a matrix-ESS) by Maynard Smith (1982). Furthermore, it is well-known a matrix-ESS is locally asymptotically stable for the replicator equation on a finite trait space.

Specifically, they showed the game with the continuum of traits in S is strictly dominance solvable (see Definition 5 below) to the trait 0. By this process, each trait $x \in S$ that is strictly dominated by another trait $y \in S$ is eliminated and then each remaining trait that is strictly dominated (in the reduced game with the resultant trait space) by another remaining trait is eliminated, etc. If every trait except 0 is eventually eliminated by this countable process, standard techniques extended from finite trait space (Samuelson and Zhang, 1992) show δ_0 is attractive for the above initial distributions with respect to the replicator equation under this iterated elimination of strictly dominated strategies (Cressman and Hofbauer, 2005) (see also Heifetz et al. (2003) and the proof of Theorem 10 below).

In this section, we extend this argument to a multi-dimensional setting, starting with the concept of dominance solvability similar to that introduced by Moulin (1984).

Definition 5. The game with compact trait space S is strictly dominance solvable to $x^* \in S$ if there is a countable nested sequence of closed subsets S_i in S with $S_{i+1} \subset S_i$ and $S_0 = S$ satisfying

- (i) for every $i \geq 0$ and every $x \in \overline{S_i \setminus S_{i+1}}$, there exists a $y \in S_i$ such that $\pi(y, z) > \pi(x, z)$ for all $z \in S_i$
- (ii) $\bigcap_{i=0}^{\infty} S_i = \{x^*\}$.

Theorem 10. If the game with compact trait space $S \subset \mathbf{R}^n$ is strictly dominance solvable to $x^* \in S$, then P_t converges to δ_{x^*} in the weak topology for each initial distribution P_0 with full support S .

The most straightforward application of this theory to our multi-dimensional setting is through the following theorem when B is symmetric (and the trait space is parameterized so that the payoff function has form (5) up to quadratic terms). A set $S \subset \mathbf{R}^n$ is called *star-shaped about 0* if it contains the line segment from 0 to x for every $x \in S$. Hence for $n = 1$, S is an interval containing 0. Let Q^* be the set of all probability measures in $\Delta(S)$ whose support is star-shaped about 0.

Theorem 11. Suppose B is symmetric. The games restricted to all compact trait spaces $S \subset \mathbf{R}^n$ that are star-shaped about 0 and sufficiently close to 0 are strictly dominance solvable if and only if $\|B\| < 2$.²¹ Furthermore, if $\|B\| < 2$, then δ_0 is neighborhood attractive for the replicator equation (1) for Q^* .

Notice that 0 is a CSS if $\|B\| < 2$ since we then have

$$x \cdot (B - 2I)x < \|B\| \|x\|^2 - 2\|x\|^2 < 0$$

²¹Here $\|B\|$ is the operator norm of B (i.e. $\|B\| \equiv \sup_{\|x\|=1} \|Bx\|$ where $\|x\|$ is the Euclidean length of x). We ignore the degenerate possibility $\|B\| = 2$. Strict dominance solvability for all games with trait space S sufficiently close to x^* is closely related to the concept of *locally strictly dominance solvable* defined by Moulin (1984).

for nonzero x (i.e. $B - 2I$ is negative definite). The condition $\|B\| < 2$ also has an interesting connection to the Cournot tatonnement process of Moulin (1984). To see this clearly, let us ignore the non-quadratic terms in Eq. (5). For our single-species model, the Cournot process is the sequence of best replies $y_{i+1} \equiv \arg \max \{\pi(x, y_i) : x \in \mathbf{R}^n\}$ which is given by $y_{i+1} = \frac{1}{2}By_i$. This discrete-time tatonnement process converges to 0 if and only if the eigenvalues of $\frac{1}{2}B$ all have modulus less than 1. For symmetric B , this is equivalent to $\|B\| < 2$.

When $\|B\| < 2$ but B is not symmetric, the proof of Theorem 11 still shows that the games restricted to all trait spaces $S \subset \mathbf{R}^n$ that are star-shaped about 0 and sufficiently close to 0 are strictly dominance solvable if $\|B\| < 2$ (and so δ_0 is still neighborhood attractive for Q^*). However, there are other choices of B with $\|B\| > 2$ for which the associated games are strictly dominance solvable (see Section 5.3). The reason for this is that we do not need to use Euclidean distance as in the proof of Theorem 11 (where the nested sequence of trait spaces are disks). The essential inequality there is that, for nonzero y_0 ,

$$y_0 \cdot Bz < 2y_0 \cdot y_0 \tag{18}$$

for all z in the disk $\{z | z \cdot z \leq y_0 \cdot y_0\}$. These regions can be replaced by others that are compact and star-shaped. For instance, if D is a positive definite symmetric matrix, we will have strict dominance solvability if $y_0 \cdot Bz < 2y_0 \cdot y_0$ for all z in the disk $z \in \{z | z \cdot Dz \leq y_0 \cdot Dy_0\}$. Since Eq. (18) is linear in z , we can restrict to the boundary $\{z | z \cdot Dz = y_0 \cdot Dy_0\}$. The following theorem then generalizes Theorem 11 to non-symmetric B .

Theorem 12. Suppose there exists a positive definite symmetric matrix D such that, for all $y \neq 0$,

$$(y \cdot Dy)(B^T y \cdot D^{-1} B^T y) < 4(y \cdot y)^2 \tag{19}$$

Then the games restricted to all compact trait spaces $S \subset \mathbf{R}^n$ that are star-shaped about 0 and sufficiently close to 0 are strictly dominance solvable and δ_0 is neighborhood attractive for Q^* in the replicator equation (1). If δ_0 is neighborhood attractive in this sense, then 0 is CSS.

Remark. If D is chosen as a (positive) multiple of the identity matrix in Theorem 12, then Eq. (19) is equivalent to $\|B^T y\| < 2\|y\|$. Since $\|B\| = \|B^T\|$, Theorem 12 shows that the statement of Theorem 11 is valid when $\|B\| < 2$ and B is non-symmetric.

Conversely, if B is symmetric with $\|B\| > 2$, then $B^T y = \lambda y$ for some y and $|\lambda| > 2$. Since $(y \cdot Dy)(y \cdot D^{-1} y) \geq (y \cdot y)^2$ for all positive definite symmetric matrices D , we have $(y \cdot Dy)(B^T y \cdot D^{-1} B^T y) > 4(y \cdot y)^2$ and so Theorem 12 does not expand the set of payoff functions with B symmetric for which we have a proof that δ_0 is neighborhood attractive.

5.3. Two-dimensional trait space

Using Theorems 11 and 12 on dominance solvability, we are able to determine a large class of 2×2 matrices B for which δ_0 is neighborhood attractive in the replicator equation (1) for Q^* . For this, we use the fact that every 2×2 matrix is orthogonally similar to a matrix of the form

$$B = \begin{bmatrix} a & b \\ c & a \end{bmatrix}. \tag{20}$$

That is, there is a rotation O such that $O^T B O$ has this form. Note that such a transformation does not affect the form of the fitness function (2) nor the symmetry of B .

Theorem 13. *If B is given by Eq. (20), then the games restricted to all compact trait spaces $S \subset \mathbf{R}^2$ that are star-shaped about 0 and sufficiently close to 0 are strictly dominance solvable if*

$$|b| + |c| < 2(2 - |a|). \tag{21}$$

In this case, δ_0 is neighborhood attractive for Q^ .*

If B is symmetric, Eq. (21) becomes $|a| + |b| < 2$ which is the condition of Theorem 11 since $\|B\| = |a| + |b| < 2$. On the other hand, by Theorem 13, there are non-symmetric B 's for which δ_0 is neighborhood attractive but $\|B\| > 2$. For instance, δ_0 is neighborhood attractive for $B = \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix}$ if $|b| < 4$ but $\|\begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix}\| = |b|$. Finally, it is straightforward to confirm that condition (21) implies the eigenvalues $a \pm \sqrt{bc}$ of B satisfy $|a \pm \sqrt{bc}| < 2$. Thus the best reply Cournot tatonnement process (Moulin, 1984) again converges to zero although it is no longer true that $\|x_{t+1}\| < \|x_t\|$.

5.3.1. Summary for two-dimensional trait space

Take $\pi(x, z) = -x \cdot x + x \cdot Bz$ and $B = \begin{bmatrix} a & b \\ c & a \end{bmatrix}$.

For B symmetric, 0 is

- (i) NIS if and only if $a < 1$ and $|b| < |1 - a|$,
- (ii) CSS if and only if $a < 2$ and $|b| < |2 - a|$,
- (iii) dominance solvable if and only if $|a| + |b| < 2$.

For B non-symmetric, 0 is

- (i) NIS if and only if $a < 1$ and $|b + c| < 2|1 - a|$,
- (ii) CSS if and only if $a < 2$ and $|b + c| < 2|2 - a|$,
- (iii) dominance solvable if $|b| + |c| < 2(2 - |a|)$.

Condition (21) is equivalent to strict dominance solvability when B is non-symmetric and $bc \geq 0$ (i.e. b and c have the same sign). To see this, Theorem 11 applied to the one-dimensional trait space $S = \{t(1, 1) : |t| \leq 1\}$ yields the necessary condition $|2a + b + c| < 4$ for strict dominance solvability and to the one-dimensional trait space $S = \{t(1, -1) : |t| \leq 1\}$ yields $|2a - b - c| < 4$. Thus, $|2a| + |b + c| < 4$. On the other hand, $\|B\| = \sqrt{1 + b^2}$ for $B = \begin{bmatrix} 1 & b \\ -b & 1 \end{bmatrix}$ and so the game is strictly dominance solvable if $|b| < \sqrt{3}$ even though $|2a| + |b| + |c|$ may be arbitrarily close to $2 + 2\sqrt{3} > 4$.

In fact, the exact condition for strict dominance solvability to 0 is unknown for non-symmetric B . It is also an open question whether the CSS condition is sufficient for neighborhood attractivity of δ_0 for Q^* .

6. Discussion

As stated in the Introduction, we feel the adaptive dynamics model to predict stability of monomorphisms by emphasizing the evolution of the population mean strategy misses the effects of the spread of the distribution of individual behaviors. The replicator equation on a continuous trait space is our preferred method to include these effects. The basic issue we consider is then whether the static CSS and NIS concepts for monomorphic stability of one-dimensional adaptive dynamics predict stability of the replicator equation when generalized to multi-dimensions. An initial obstacle to analysing this issue is that universally accepted static extensions are not agreed upon in the adaptive dynamics approach since stability of the canonical equation now depends on the relative rates mutations occur in different directions (in technical terms, on the mutations' covariance matrix).

One assumption is that the covariance matrix will evolve very slowly (if at all) and so can be taken as essentially constant (Vincent et al., 1993), a method that has also been used effectively much earlier in the matrix game model (Hines, 1980a). With arbitrary (but fixed) covariance, monomorphic stability with respect to both the canonical equation and to the potential evolution of dimorphisms leads to the CSS conditions in each direction through the monomorphism (Meszéna et al., 2001). We take this as our multi-dimensional CSS concept. On the other hand, if the relative rates of mutation are not constant but can change at different points along the evolutionary path to have their most extreme effect (Leimar, 2001, 2005), much stronger stability conditions than being a CSS in each direction are needed in the adaptive dynamics approach (see Section 4).

In light of the above discussion, the analytic results of Section 3 are quite surprising. By Theorem 3 there, when individual behaviors are initially normally distributed and fitnesses are approximated by their second-order Taylor expansions about a monomorphism, then even the weaker CSS conditions are too strong (see Theorem 3 for the precise statement) to characterize stability since the distribution's covariance evolves slowly to having equal effect in all directions. Although this result raises legitimate concerns about current adaptive dynamics approaches for multi-dimensional trait space, the presence of strategies at the tails of the normal distribution does not match the usual assumption that mutations only occur near the monomorphic equilibrium.

For this reason, Section 5 concentrates on behavioral distributions with compact support contained in a small neighborhood of a monomorphism m^* .²² In this setting

²²For technical reasons, this support must also include m^* .

and with no more restrictions on the support of the initial distribution, we show in Section 5.1 that the multi-dimensional NIS is the most relevant concept for stability of δ_{m^*} under the replicator equation (Theorems 7–9), generalizing results of Eshel and Sansone (2003); Cressman (2005). When distributions are restricted to those whose supports are compact intervals in each direction from m^* , we show in Section 5.2 the relevance of the CSS conditions by clarifying the relationship between CSS and strategy dominance in the multi-dimensional model. As explained there, stability of the measure dynamics via strategy domination corresponds to stability of the Cournot tatonnement process with a continuum of strategies (Moulin, 1984), a discrete-time dynamic whereby rational decision makers choose the optimal strategy in the next time period given current population behavior. This connection continues the long tradition of classical game-theoretic methods providing valuable insight into the eventual outcome of behavioral evolution under the replicator equation (and vice versa).

Finally, Section 5.3 summarizes how our results apply to two-dimensional trait spaces, an important special case that highlights the added difficulties that arise when trait space has an extra degree of freedom compared to the analysis of Cressman and Hofbauer (2005) where trait space is one dimensional.

Appendix

Proof of Theorem 1. The proof that the class of normal distributions is invariant under replicator dynamics is in Oechssler and Riedel (2001) for $n = 1$. A similar method works for many dimensions and it establishes the system of ordinary differential equations (6) and (7) at the same time. In the following, a different proof by using moment generating functions is given.

First, we provide some intuition for the proof of Eq. (6). Assume that all P_t are normally distributed with mean $m(t)$ and diagonal covariance matrix with entries $V_i(t)$. Note that $\pi(x, P_t) = -x \cdot x + x \cdot Bm(t)$ and $\pi(P_t, P_t) = -m(t) \cdot m(t) - V(t) \cdot 1 + x \cdot Bm(t)$, where 1 denotes the summing vector $(1, 1, \dots, 1)$. Then the differential equation for the i th component of the mean is (omitting the time variable t)

$$\begin{aligned} m'_i &= \int_{\mathbf{R}^n} x_i(\pi(x, P) - \pi(P, P))P(dx) \\ &= \int_{\mathbf{R}^n} \left(-x_i^3 + x_i \left(\sum_{k=1}^n (x_i - m_i) B_{ik} m_k \right) \right) P(dx) \\ &\quad + \sum_{j \neq i} \int_{\mathbf{R}^n} -x_i \left(x_j^2 + \left(\sum_{k=1}^n (x_j - m_j) B_{jk} m_k \right) \right) P(dx) \\ &\quad - m_i(-m \cdot m - V \cdot 1). \end{aligned}$$

The third moment of a normal random variable is $\int x_i^3 P(dx) = m_i^3 + 3m_i V_i$. The covariance of x_i and $x_i - m_i$ is equal to the variance V_i . As we have assumed that the

components are uncorrelated, $\int x_i(x_j - m_j)P(dx) = 0$ for $i \neq j$. Substitution of these results into the last equation for m'_i yields

$$m'_i = -2m_i V_i + V_i \sum_{k=1}^n B_{ik} m_k.$$

That is, in vector notation, we have Eq. (6).

When the covariance matrix is not diagonal, the following proof uses moment generating functions. As this method has potential use in other contexts as well, we start with a general exposition here.

Take a probability measure P with mean m and covariance matrix C and a vector $\lambda \in \mathbf{R}^n$. Define the Laplace transform

$$L(\lambda; P) = \int \exp(-\lambda^T x) P(dx),$$

and its logarithm

$$M(\lambda; P) = \log L(\lambda; P).$$

L can be extended to finite signed measures in a straightforward way. It is well known (and follows immediately through differentiation under the integral) that

$$\left. \frac{\partial}{\partial \lambda_i} M(\lambda; P) \right|_{\lambda=0} = -m_i$$

and

$$\left. \frac{\partial^2}{\partial \lambda_i \partial \lambda_j} M(\lambda; P) \right|_{\lambda=0} = C_{ij}.$$

Thus, M generates the mean through the gradient and the covariance matrix through its Hessian. Therefore, we can obtain differential equations for the mean and the covariance matrix by differentiating the moment generating function M . Normal distributions $N(m, C)$ are characterized by $M(\lambda, N(m, C)) = -\lambda^T m + \frac{1}{2} \lambda^T C \lambda$.

Let $P(t)$ be a trajectory of replicator dynamics in the following. It is useful to associate with the measures $P(t)$ the probability measures $P_\lambda(t)$ as given by

$$P_\lambda(t)(\Gamma) = \frac{1}{L(\lambda; P(t))} \int_{\Gamma} \exp(-\lambda^T x) P(t)(dx).$$

Note that

$$\frac{d}{dt} M(\lambda; P(t)) = \frac{L(\lambda; P'(t))}{L(\lambda; P(t))}.$$

By definition of replicator dynamics

$$\begin{aligned} \frac{d}{dt} M(\lambda; P(t)) &= \frac{1}{L(\lambda; P(t))} \int \exp(-\lambda^T x) [\pi(x, P(t)) \\ &\quad - \pi(P(t), P(t))] P(t)(dx) \\ &= \frac{1}{L(\lambda; P(t))} \int \exp(-\lambda^T x) \pi(x, P(t)) P(t)(dx) \\ &\quad - \pi(P(t), P(t)) \\ &= \pi(P_\lambda(t), P(t)) - \pi(P(t), P(t)). \end{aligned}$$

From this, we get the following relations for the mean and the covariance matrix:

$$m'_i(t) = -\frac{\partial}{\partial \lambda_i} \pi(P_\lambda(t), P(t)) \Big|_{\lambda=0},$$

$$C'_{ij}(t) = \frac{\partial^2}{\partial \lambda_i \partial \lambda_j} \pi(P_\lambda(t), P(t)) \Big|_{\lambda=0}.$$

We will now apply these relationships to the case of normal distributions. As is well known (and can be seen via completing a square), if $P = N(m, C)$, then $P_\lambda = N(m - C\lambda, C)$. Thus, for the quadratic payoff function $\pi(x, y) = x \cdot Ax + x \cdot By$,

$$\pi(P_\lambda, P) = (m - C\lambda)^T A(m - C\lambda) + \sum_{i,j=1}^n A_{ij} C_{ij} + (m - C\lambda)^T Bm.$$

Since this is quadratic in λ , the set of normal distributions is invariant, and by comparing with $(d/dt)M(\lambda; P(t)) = -\lambda^T m'(t) + \frac{1}{2} \lambda^T C'(t) \lambda$, we get the desired differential equations

$$m'(t) = C(t)(2A + B)m(t), \tag{22}$$

$$C'(t) = 2C(t)AC(t) \tag{23}$$

which reduces in the special case $A = -I$ to Eqs. (6) and (7). \square

Proof of Theorem 3. As a first step, we perform an orthogonal transformation as in the analysis of the dynamics of the covariance matrix. Let $C = C(0)$ be the initial covariance matrix, and choose an orthogonal matrix O and a diagonal matrix D with decreasing entries $D_{11} \geq \dots \geq D_{nn} > 0$ such that $OCO^T = D$. We know from the discussion after Theorem 2 that $C(t) = O^T D(t) O$ for $D(t)$ diagonal with $D_{ii}(t) = 1/(D_{ii}^{-1} + 2t)$. From this, it is straightforward to show that $\min_{i,j} D_{ii}(t)/D_{jj}(t) = D_{nn}(t)/D_{11}(t)$ and that $D_{nn}(t)/D_{11}(t) \geq D_{nn}/D_{11}$. Hence, the set of covariance matrices with bounded ratios of eigenvalues $\min_{i,j} D_{ii}/D_{jj} = D_{nn}/D_{11} \geq \varepsilon > 0$ is forward invariant under Eq. (7).

Define the transformed mean vectors $n(t) = Om(t)$. As O is orthogonal, $n(t)$ and $m(t)$ have the same distance to zero and it suffices to study the stability behavior of $n(t)$. $n(t)$ solves the differential equation

$$n'(t) = Om'(t) = OC(t)(B - 2I)m(t) = OO^T D(t)O(B - 2I)m(t) = D(t)O(B - 2I)O^T n(t).$$

Let $\Gamma = O(B - 2I)O^T$. Γ has the same eigenvalues as $B - 2I$ since they are similar matrices.

In the second step, we perform a suitable time change by setting $s(t) = (e^{2t} - 1)/2D_{11}$. Let $\tilde{n}(t) = n(s(t))$. Then,

$$\tilde{n}'(t) = \Gamma \tilde{n}(t) - \tilde{D}(t) \Gamma \tilde{n}(t), \tag{24}$$

where $\tilde{D}(t)$ is diagonal with entries

$$\tilde{D}_{ii}(t) = \frac{a_i}{a_i + e^{2t}} \quad \text{and} \quad a_i = \frac{D_{11}}{D_{ii}} - 1.$$

The system of differential equations (24) is asymptotically autonomous with exponentially decreasing remainder term. Therefore, one can invoke standard results (see, for example, Coddington and Levinson, 1955, Chapter 3, Problem 35) to conclude that the eigenvalues of Γ (i.e., those of $B - 2I$) determine the asymptotic behavior of $m(t)$ if the positive definite covariance matrix $C(0)$ is fixed. In particular, if one eigenvalue of Γ has positive real part, the mean diverges for most initial conditions (Part 2) and $m = 0$ is asymptotically stable if every eigenvalue of Γ has negative real part (first statement of Part 1).

For the second statement of Part 1, we do vary $C(0)$ and so need to work a little more using our uniform estimate in the initial covariances. By our assumption on the eigenvalues of $C(0)$, we have $a_i/(a_i + e^{2t}) \leq b/(b + e^{2t})$ with $b = 1/\varepsilon - 1$. By the variations of constants formula, the solution of Eq. (24) can be written as

$$\tilde{n}(t) = \exp(\Gamma t) \tilde{n}(0) - \int_0^t \exp(\Gamma(t-s)) \tilde{D}(s) \Gamma \tilde{n}(s) ds.$$

As Γ is stable, there exist constants $K, \eta > 0$ such that $\|\exp(\Gamma t)\| \leq K \exp(-\eta t)$ for all t . It follows that

$$\|\tilde{n}(t)\| \leq K \exp(-\eta t) \|\tilde{n}(0)\| + \int_0^t K \exp(-\eta(t-s)) \frac{b}{b + e^{2s}} \|\Gamma\| \|\tilde{n}(s)\| ds$$

or

$$\exp(\eta t) \|\tilde{n}(t)\| \leq K \|\tilde{n}(0)\| + \int_0^t K \frac{b}{b + e^{2s}} \|\Gamma\| \exp(\eta s) \|\tilde{n}(s)\| ds.$$

By applying Gronwall's lemma (Coddington and Levinson, 1955, Chapter 1, Problem 1) to the function $f(t) = \exp(\eta t) \|\tilde{n}(t)\|$, we obtain

$$f(t) \leq K \|\tilde{n}(0)\| \exp\left(K \|\Gamma\| \int_0^t \frac{b}{b + e^{2s}} ds\right) \leq K \|\tilde{n}(0)\| \exp\left(K \|\Gamma\| \int_0^\infty \frac{b}{b + e^{2s}} ds\right) = K \|\tilde{n}(0)\| (1 + b)^{K \|\Gamma\|/2}.$$

The constant $L = K(1 + b)^{K \|\Gamma\|/2}$ does not depend on the initial conditions $C(0)$ and $m(0)$. Thus, we get the estimate

$$\|\tilde{n}(t)\| \leq L \exp(-\eta t) \|\tilde{n}(0)\|$$

which establishes asymptotic stability in Part 1.

At the same time, this estimate shows that δ_0 attracts all trajectories that start at some $N(m, C)$ with a positive definite covariance matrix. However, since the number $b = 1/\varepsilon - 1$ tends to infinity as ε tends to zero, our argument does not establish asymptotic stability for the set of all positive definite initial covariance matrices (which requires stronger assumptions on B). \square

Proof of Theorem 4. Let $\tilde{B} = B - 2I$ and $P = xx^T$ be a symmetric rank one matrix (which arises from any initial normal distribution concentrated on the line through the nonzero vector $x \in \mathbf{R}^n$, i.e. the kernel of $C(0)$ is perpendicular to x). Then stability of Eq. (12) implies stability of the matrix $P\tilde{B} = xx^T\tilde{B}$ which in turn implies $0 \geq \text{tr}(P\tilde{B}) = \text{tr}(xx^T\tilde{B}) = \text{tr}(x^T\tilde{B}x) = x^T\tilde{B}x$.²³ Since this is true for all x , $\tilde{B} = B - 2I$ is negative semidefinite. \square

The following Lemma generalizes a familiar statement from finite games, a part of the *folk theorem of evolutionary game theory*, see Cressman (2003); Hofbauer and Sigmund (1998), to infinite games.

Lemma A.1. For a compact metric space S , let the initial P_0 have full support S and $P_t \rightarrow P$ in the weak topology. Then P is a Nash equilibrium.

Proof. Denote $\sigma(x, P) = \pi(x, P) - \pi(P, P)$. If $P_t \rightarrow P$, then $\sigma(s, P_t) \rightarrow \sigma(s, P)$ for each s . Suppose that $\sigma(s, P) > 0$ for some $s \in S$. By continuity, there is a compact set $A \subseteq S$ (with nonempty interior, and hence $P_0(A) > 0$) such that $\sigma(s, P) \geq 2\varepsilon > 0$ for all $s \in A$. Thus, for each $s \in A$, there is a $t_s > 0$ such that $\frac{1}{t} \int_0^t \sigma(s, P_\tau) d\tau \geq \varepsilon > 0$ for all $t \geq t_s$.

By Lemma 2 in Bomze (1991), $dP_t/dP_0(x) = \exp(\int_0^t \sigma(x, P_\tau) d\tau) \geq e^{\varepsilon t}$ for $t \geq t_x$. Hence $P_t(A) = \int_A dP_t/dP_0(x)P_0(dx) \rightarrow \infty$ as $t \rightarrow \infty$ by Fatou's lemma (Royden, 1988). Since $P_t(A) \leq 1$, this is a contradiction. \square

Proof of Theorem 8. The mapping $\sigma : (x, P) \rightarrow \pi(x, P) - \pi(P, P)$ is jointly continuous in x (Euclidean topology) and P (weak topology). As δ_0 is neighborhood superior, we have $\sigma(0, P) > 0$ for all $P \neq \delta_0$ whose support is within an ε_0 ball of 0. If δ_0 is not a limit point in the weak topology of an initial P_0 with such support, then there is an open neighborhood of δ_0 that includes no P_t for all t sufficiently large. The set of all P outside this neighborhood with support within the ε_0 ball is compact in the weak topology. By continuity in P , we have $\sigma(0, P) \geq \kappa > 0$ for all such P and some κ . By continuity in x , we can find some $\varepsilon_1 > 0$ such that we have $\sigma(x, P) \geq \kappa/2 > 0$ for all $|x| \leq \varepsilon_1$ and all such P . But this implies

$$\frac{P'(t)(U)}{P(t)(U)} \geq \frac{\kappa}{2},$$

for the ball $U = \{x \in S : |x| \leq \varepsilon_1\}$. Then $P(t)(U) \uparrow \infty$, which is a contradiction. \square

Proof of Theorem 9. From Oechssler and Riedel (2002, Theorem 1), we know that mean fitness $\pi(P, P)$ is a Lyapunov function, so that $t \mapsto \pi(P_t, P_t)$ is monotonically increasing. For any initial P_0 with compact support close to 0 and containing 0, Theorem 8 implies that $\pi(P_t, P_t) \rightarrow \pi(0, 0)$ as $t \rightarrow \infty$. Neighborhood superiority of δ_0 implies 0 is a local NE and $\pi(0, 0) > \pi(P, P)$ for all $P \neq \delta_0$ with support close to 0 (cf. the remark in Cressman and

Hofbauer, 2005, p. 53). Hence δ_0 is the only ω -limit point of P_t and hence $P_t \rightarrow \delta_0$.

This completes the proof that δ_0 is neighborhood attractive if δ_0 is neighborhood superior. The converse follows from the stability analysis of Eq. (15) and Theorem 7. \square

Proof of Theorem 10. It is sufficient to show by induction on i that $\lim_{t \rightarrow \infty} P_t(S \setminus S_i) = 0$ for all $i \geq 1$. Given $x_0 \in S \setminus S_1$, there exists a $y_0 \in S$ such that $\pi(y_0, z) > \pi(x_0, z)$ for all $z \in S$. By continuity of π , there are open neighborhoods $U(x_0)$ and $U(y_0)$ of x_0 and y_0 , respectively so that

$$\pi(y, z) - \pi(x, z) \geq K > 0$$

for all $x \in U(x_0), y \in U(y_0)$ and $z \in S$. Since $\text{supp}(P_t) = S$, both $P_t(U(x_0))$ and $P_t(U(y_0))$ are positive. From Eq. (1), an application of the quotient rule yields

$$\begin{aligned} \frac{d}{dt} \left(\frac{P_t(U(y_0))}{P_t(U(x_0))} \right) &= \frac{1}{P_t(U(x_0))^2} \int_S \left[\int_{U(x_0)} \int_{U(y_0)} (\pi(y, z) \right. \\ &\quad \left. - \pi(x, z)) P_t(dx) P_t(dy) \right] P_t(dz) \\ &\geq K \frac{P_t(U(y_0))}{P_t(U(x_0))}. \end{aligned}$$

Thus $\lim_{t \rightarrow \infty} P_t(U(y_0))/P_t(U(x_0)) = \infty$ and, in particular, $P_t(U(x_0))$ converges to 0. Since $S \setminus S_1$ is compact, it is covered by finitely many $U(x_0)$ and so $\lim_{t \rightarrow \infty} P_t(S \setminus S_1) = 0$.

Now assume $\lim_{t \rightarrow \infty} P_t(S \setminus S_i) = 0$. As above, for every $x_0 \in S_i \setminus S_{i+1}$, there exists a $y_0 \in S_i$ such that $\pi(y_0, z) > \pi(x_0, z)$ for all $z \in S_i$. By continuity of π , there are open neighborhoods $U(x_0)$ and $U(y_0)$ of x_0 and y_0 , respectively so that

$$\pi(y, z) - \pi(x, z) \geq K > 0$$

for all $x \in U(x_0), y \in U(y_0)$ and $z \in S_i$. Let $k \equiv \max \{|\pi(y, z) - \pi(x, z)| : x, y, z \in S\}$. Then

$$\begin{aligned} \frac{d}{dt} \left(\frac{P_t(U(y_0))}{P_t(U(x_0))} \right) &= \frac{1}{P_t(U(x_0))^2} \int_{S \setminus S_i} \left[\int_{U(x_0)} \int_{U(y_0)} (\pi(y, z) \right. \\ &\quad \left. - \pi(x, z)) P_t(dx) P_t(dy) \right] P_t(dz) \\ &\quad + \frac{1}{P_t(U(x_0))^2} \int_{S_i} \left[\int_{U(x_0)} \int_{U(y_0)} (\pi(y, z) \right. \\ &\quad \left. - \pi(x, z)) P_t(dx) P_t(dy) \right] P_t(dz) \\ &\geq [-k P_t(S \setminus S_i) + K P_t(S_i)] \frac{P_t(U(y_0))}{P_t(U(x_0))} \\ &> 0 \end{aligned}$$

for t sufficiently large since $\lim_{t \rightarrow \infty} P_t(S \setminus S_i) = 0$ and $\lim_{t \rightarrow \infty} P_t(S_i) = 1$. The proof continues as above to yield $\lim_{t \rightarrow \infty} P_t(S_i \setminus S_{i+1}) = 0$ and so $\lim_{t \rightarrow \infty} P_t(S \setminus S_{i+1}) = 0 = \lim_{t \rightarrow \infty} P_t(S \setminus S_i) + \lim_{t \rightarrow \infty} P_t(S_i \setminus S_{i+1}) = 0$ for all $i \geq 1$.

²³Here $\text{tr}(P\tilde{B})$ is the trace of $P\tilde{B}$.

Therefore, $\lim_{t \rightarrow \infty} P_t(S \setminus S_t) = 0$ and so P_t converges weakly to δ_0 . \square

Proof of Theorem 11. Suppose that $\|B\| > 2$. By the symmetry of B , there exists a (nonzero) eigenvector x_0 such that its corresponding eigenvalue is $\pm\|B\|$. Consider the star-shaped compact trait spaces $S \subset \mathbf{R}^n$ of the form $\{cx_0 : |c| \leq \varepsilon\}$ for some $\varepsilon > 0$. That is, S is the line segment through the origin joining $-\varepsilon x_0$ to εx_0 . We claim no $x \in S$ can be eliminated through strict domination by another $y \in S$ if ε is sufficiently small. To see this, we parameterize our trait space so that it is a subset of \mathbf{R} (i.e. parameterized by a scalar x) with $x_0 = 1$. We then have $Bx_0 = bx_0$ where $|b| > 2$. From Eq. (4), we have

$$\pi(x, z) - \pi(y, z) = (y - x)[x + y - bz + h.o.t.]$$

where the higher-order terms are at least of degree two in the variables x, y, z . Since $|x| \leq \varepsilon, |y| \leq \varepsilon$ and $|b| > 2$, when ε is sufficiently small the expression $x + y - bz + h.o.t.$ is positive for all $x, y \in S$ by either choosing z as $-\varepsilon$ or ε appropriately and negative for all $x, y \in S$ by choosing the alternate z . That is, for all $x, y \in S$, there is a $z \in S$ with $\pi(x, z) - \pi(y, z) \geq 0$ and so $x \in S$ cannot be strictly dominated by any $y \in S$.

Now suppose $\|B\| < 2$. Let $d \equiv \max_{z \in S} \|z\|$. Suppose $x_0 \in S$ and $\|x_0\|$ is close to d . From Eq. (4) with $y_0 = (1 - \varepsilon)x_0 \in S$,

$$\begin{aligned} \pi(y_0, z) - \pi(x_0, z) &= \varepsilon[(2 - \varepsilon)\|x_0\|^2 - x_0 \cdot Bz + h.o.t.] \\ &> \frac{\varepsilon}{2}[2\|x_0\|^2 - \|x_0\|\|Bz\|] \\ &> \frac{\varepsilon}{4} \frac{\|x_0\|^2}{2} [2 - \|B\|] > 0 \end{aligned}$$

for all $z \in S$. That is, y_0 strictly dominates x_0 . By continuity of π , y_0 strictly dominates x for all x near x_0 . Define $A(\alpha, \beta)$ with $0 \leq \alpha < \beta < d$ as those elements of S in the annulus $\{y \in S : d - \beta \leq \|y\| \leq d - \alpha\}$. Each $A(\alpha, \beta)$ is compact and non-empty by the definition of d and the fact S is star-shaped. Thus, for some $\delta > 0$, each $x \in A(0, \delta)$ is strictly dominated and so we can take $S_1 = \{y \in S : \|y\| \leq d - \delta\}$. We now iterate this argument and obtain our sequence S_i of nested sets as the intersection of S with a disk centered at 0 whose radius tends to 0 as $i \rightarrow \infty$.

We next show δ_0 is attractive if $\|B\| < 2$. By the argument used in the preceding paragraph, we can apply Theorem 10 and conclude that P_t converges to δ_0 in the weak topology whenever P_0 has star-shaped support sufficiently close to 0. That is, δ_0 is neighborhood attractive for Q^* . \square

The following lemma is used in the proof of Theorem 12 below.

Lemma A.2. *If D is a positive definite symmetric matrix, the maximum of $y_0 \cdot Bz$ subject to $\{z | z \cdot Dz = y_0 \cdot Dy_0\}$ is $\sqrt{y_0 \cdot Dy_0} \sqrt{B^T y_0 \cdot D^{-1} B^T y_0}$ and this occurs at $z_0 = \sqrt{y_0 \cdot Dy_0} / \sqrt{B^T y_0 \cdot D^{-1} B^T y_0} D^{-1} B^T y_0$ if $B^T y_0 \neq 0$.*

Proof. If $B^T y_0 = 0$, there is nothing to prove so assume $B^T y_0 \neq 0$. The result may be shown by using Lagrange multipliers or by the following algebraic method. Let \sqrt{D} be the positive definite square root of D . Then

$$\begin{aligned} (y_0 \cdot Bz)^2 &= (y_0 \cdot B\sqrt{D}^{-1}\sqrt{D}z)^2 \\ &= (\sqrt{D}^{-1}B^T y_0 \cdot \sqrt{D}z)^2 \\ &\leq (\sqrt{D}^{-1}B^T y_0 \cdot \sqrt{D}^{-1}B^T y_0) (\sqrt{D}z \cdot \sqrt{D}z) \\ &= (B^T y_0 \cdot D^{-1}B^T y_0) (z \cdot Dz) \\ &= (B^T y_0 \cdot D^{-1}B^T y_0) (y_0 \cdot Dy_0). \end{aligned}$$

It is straightforward to verify z_0 satisfies the requirements. \square

Proof of Theorem 12. By Lemma A.2, a straightforward generalization of the relevant parts of the proof of Theorem 11 yields the first result. Now suppose 0 is not CSS. Then $B - 2I$ is not negative definite and so there is a nonzero x such that $x \cdot (B - 2I)x > 0$. Take S as the one-dimensional bounded interval in the direction x that includes 0 in its interior. Note that S is star-shaped. By Cressman and Hofbauer (2005), δ_0 is not neighborhood attractive for Q^* . \square

Proof of Theorem 13. Suppose $|b| + |c| < 2(2 - |a|)$. Take the matrix $D = \begin{bmatrix} |a|+|c| & 0 \\ 0 & |a|+|b| \end{bmatrix}$ in (19).²⁴ The left-hand side of Eq. (19) is then

$$\begin{aligned} &\left[\frac{(ay_1 + cy_2)^2}{|a| + |c|} + \frac{(by_1 + ay_2)^2}{|a| + |b|} \right] \\ &\quad \times [(|a| + |c|)(y_1)^2 + (|a| + |b|)(y_2)^2] \end{aligned}$$

which is less than or equal to

$$\begin{aligned} &\left[\frac{(|ay_1| + |cy_2|)^2}{|a| + |c|} + \frac{(|by_1| + |ay_2|)^2}{|a| + |b|} \right] \\ &\quad \times [(|a| + |c|)(y_1)^2 + (|a| + |b|)(y_2)^2]. \end{aligned}$$

By considering the cases where $|y_1| \leq |y_2|$ and $|y_1| \geq |y_2|$ separately, it is straightforward to show that this last expression is increasing in $|c|$ for fixed y_1, y_2 and $|b|$. Thus we can replace $|c|$ by $2(2 - |a|) - |b|$. Furthermore, for fixed $|b|$ and $y_2 = ky_1$ for $k > 0$, the right-hand side of Eq. (19) minus this last expression has an absolute maximum when $k = 1$. For $k = 1$, the value is zero and so the games are strictly dominance solvable by Theorem 12. \square

References

- Abrams, P., 2001. Modelling the adaptive dynamics of traits involved in inter- and intraspecific interactions: an assessment of three methods. *Ecol. Lett.* 4, 166–175.
- Alós-Ferrer, C., Ania, A.B., 2001. Local equilibria in economic games. *Econ. Lett.* 70, 165–173.

²⁴If $|a| + |c| = 0$, then $B = \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix}$. Take $D = \begin{bmatrix} \varepsilon & 0 \\ 0 & |b| \end{bmatrix}$ for $\varepsilon > 0$. If $|b| < 4$, there is an $\varepsilon > 0$ such that (19) holds.

- Apaloo, J., 1997. Revisiting strategic models of evolution: the concept of neighborhood invader strategies. *Theoret. Popul. Biol.* 52, 71–77.
- Bomze, I., 1990. Dynamical aspects of evolutionary stability. *Monatsh. Math.* 110, 189–206.
- Bomze, I., 1991. Cross entropy minimization in uninhabitable states of complex populations. *J. Math. Biol.* 30, 73–87.
- Bomze, I., Pötscher, B., 1989. *Game Theoretic Foundations of Evolutionary Stability. Lecture Notes in Economics and Mathematical Systems*, vol 324. Springer, Berlin.
- Bürger, R., 2000. *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley, New York.
- Christiansen, F.B., 1991. On conditions for evolutionary stability for a continuously varying character. *Am. Nat.* 138, 37–50.
- Coddington, E.A., Levinson, N., 1955. *Theory of Ordinary Differential Equations*. McGraw-Hill, New York.
- Cressman, R., 2003. *Evolutionary Dynamics and Extensive Form Games*. MIT Press, Cambridge, MA.
- Cressman, R., 2005. Stability of the replicator equation with continuous strategy space. *Math. Soc. Sci.* 50, 127–147.
- Cressman, R., Hines, W.G.S., 1984. Correction to the appendix of three characterizations of population strategy stability. *J. Appl. Probab.* 21, 213–214.
- Cressman, R., Hofbauer, J., 2005. Measure dynamics on a one-dimensional continuous trait space: theoretical foundations for adaptive dynamics. *Theoret. Popul. Biol.* 67, 47–59.
- Dieckmann, U., Law, R., 1996. The dynamical theory of coevolution: a derivation from stochastic processes. *J. Math. Biol.* 34, 579–612.
- Doebeli, M., Dieckmann, U., 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *Am. Nat.* 156, S77–S101.
- Eshel, I., 1983. Evolutionary and continuous stability. *J. Theoret. Biol.* 103, 99–111.
- Eshel, I., Sansone, E., 2003. Evolutionary and dynamic stability in continuous population games. *J. Math. Biol.* 46, 445–459.
- Heifetz, A., Shannon, C., Spiegel, Y., 2003. What to maximize if you must, mimeo.
- Hines, W.G.S., 1980a. Strategy stability in complex populations. *J. Appl. Probab.* 17, 600–610.
- Hines, W.G.S., 1980b. Three characterizations of population strategy stability. *J. Appl. Probab.* 17, 333–340.
- Hofbauer, J., Sigmund, K., 1990. Adaptive dynamics and evolutionary stability. *Appl. Math. Lett.* 3, 75–79.
- Hofbauer, J., Sigmund, K., 1998. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge.
- Kisdi, É, Meszéna, G., 1995. Life histories with lottery competition in a stochastic environment: ESSs which do not prevail. *Theoret. Popul. Biol.* 47, 191–211.
- Leimar, O., 2001. Evolutionary change and Darwinian demons. *Selection* 2, 65–72.
- Leimar, O., 2005. Multidimensional convergence stability and the canonical adaptive dynamics. In: Dieckmann, U., Metz, J.A.J. (Eds.), *Elements of Adaptive Dynamics*. Cambridge University Press, Cambridge in press.
- Lessard, S., 1990. Evolutionary stability: one concept, several meanings. *Theoret. Popul. Biol.* 37, 159–170.
- Marrow, P., Dieckmann, U., Law, R., 1996. Evolutionary dynamics of predator–prey systems: an ecological perspective. *J. Math. Biol.* 34, 556–578.
- Maynard Smith, J., 1982. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.
- McKelvey, R., Apaloo, J., 1995. The structure and evolution of competition-organized ecological communities. *Rocky Mount. J. Math.* 25, 417–436.
- Meszéna, G., Kisdi, E., Dieckmann, U., Geritz, S.A.H., Metz, J.A.J., 2001. Evolutionary optimisation models and matrix games in the unified perspective of adaptive dynamics. *Selection* 2, 193–210.
- Metz, J.A.J., Geritz, S.A.H., Meszéna, G., Jacobs, F.J.A., van Heerwaarden, J.S., 1996. Adaptive dynamics a geometrical study of the consequences of nearly faithful reproduction. In: van Strien, S.J., Verduyn Lunel, S.M. (Eds.), *Stochastic and Spatial Structures of Dynamical Systems*. Amsterdam, North-Holland, pp. 183–231.
- Moulin, H., 1984. Dominance-solvability and Cournot stability. *Math. Soc. Sci.* 7, 83–102.
- Oechssler, J., Riedel, F., 2001. Evolutionary dynamics on infinite strategy spaces. *Econ. Theory* 17, 141–162.
- Oechssler, J., Riedel, F., 2002. On the dynamic foundation of evolutionary stability in continuous models. *J. Econ. Theory* 107, 223–252.
- Royden, H.L., 1988. *Real Analysis*, third ed. Macmillan, New York.
- Samuelson, L., Zhang, J., 1992. Evolutionary stability in asymmetric games. *J. Econ. Theory* 59, 363–391.
- Taylor, P.D., 1989. Evolutionary stability in one-parameter models under weak selection. *Theoret. Popul. Biol.* 36, 125–143.
- Taylor, P.D., Jonker, L., 1978. Evolutionarily stable strategies and game dynamics. *Math. Biosci.* 40, 145–156.
- Vincent, T.L., Cressman, R., 2000. An ESS maximum principle for matrix games. *Theoret. Popul. Biol.* 58, 173–186.
- Vincent, T.L., Cohen, Y., Brown, J.S., 1993. Evolution via strategy dynamics. *Theoret. Popul. Biol.* 44, 149–176.
- Vincent, T.L., Van, M.V., Goh, B.S., 1996. Ecological stability, evolutionary stability and the ESS maximum principle. *Evol. Ecol.* 10, 567–591.