The ESS and replicator equation in matrix games under time constraints

József Garay*
MTA-ELTE Theoretical Biology and Evolutionary Ecology Research Group
and
Department of Plant Systematics, Ecology and Theoretical Biology,
Eötvös Loránd University,
Pázmány Péter sétány 1/c, H-1117 Budapest, Hungary
and
MTA Centre for Ecological Research, Evolutionary Systems Research Group,
Klebelsberg Kuno utca 3, Tihany 8237, Hungary.
e-mail: garayj@caesar.elte.hu

Ross Cressman
Department of Mathematics
Wilfrid Laurier University
Waterloo, Ontario, Canada N2L 3C5
e-mail: rcressman@wlu.ca

Tamás F. Móri
Department of Probability Theory and Statistics,
Eötvös Loránd University,
Pázmány Péter sétány 1/c, H-1117 Budapest, Hungary
e-mail: moritamas@ludens.elte.hu

1

Tamás Varga

MTA-SZTE Analysis and Stochastics Research Group,
Bolyai Institute, University of Szeged,
Aradi vértanúk tere 1., H-6720, Szeged, Hungary
e-mail:vargata@math.u-szeged.hu

December 2, 2017

---

*Author for correspondence

Abstract

Recently, we introduced the class of matrix games under time constraints and characterized the concept of (monomorphic) evolutionarily stable strategy (ESS) in them. We are now interested in how the ESS is related to the existence and stability of equilibria for polymorphic populations. We point out that, although the ESS may no longer be a polymorphic equilibrium, there is a connection between them. Specifically, the polymorphic state at which the average strategy of the active individuals in the population is equal to the ESS is an equilibrium of the polymorphic model.

Moreover, in the case when there are only two pure strategies, a polymorphic equilibrium is locally asymptotically stable under the replicator equation for the pure-strategy polymorphic model if and only if it corresponds to an ESS. Finally, we prove that a strict Nash equilibrium is a pure-strategy ESS that is a locally asymptotically stable equilibrium of the replicator equation in n-strategy time-constrained matrix games.

Key words: evolutionary stability, monomorphic, polymorphic, replicator equation

3

1 Introduction

In ecology, the number of individuals ready to interact with the conspecifics they meet (we call these active individuals) is less than the total number of individuals in the species. For instance, other activities such as the time to handle prey (Holling 1959, Garay and Móri 2010) and the time to recover from a fight (Garay et al. 2015b; Sirot 2000) decrease the number of active individuals in a predator/consumer species. Moreover, the amount of time each of these activities take varies and may depend on the strategies (or phenotypes) used by the individuals. Consequently, in optimal foraging theory (Charnov 1976; Garay et al. 2012) and in ecological games (e.g. Broom et al. 2008, Broom and Rychtar 2013; Garay et al 2015a), activity dependent time constraints have an essential effect on the expected evolutionary outcome.

Motivated by these facts, we recently developed the theory of single-species matrix games under time constraints and characterized the concept of a (monomorphic) evolutionarily stable strategy (ESS) in them (Garay et al. 2017). This theory is briefly summarized in Section 2 below by following the static ESS approach of Maynard Smith (1982). This approach assumes that there are only two phenotypes in the population at a given time, one of which is the fixed phenotype of the resident population and the other is a rare mutant phenotype. An ESS is then a resident phenotype whose fitness is higher than that of any possible mutant (cf. Maynard Smith 1982, Garay et al. 2017). Although the static ESS concept relies implicitly upon an underlying dynamic (see Maynard Smith (1982) as well as the discussion following Definition 2.1 below), its basic intuition concentrates on the evolutionary question: What phenotype is uninvadable by an arbitrary rare mutant?

The second solution concept of evolutionary game theory (Cressman, 1992) corresponds to a stable rest point of an explicit evolutionary dynamics that models how the distribution of individual phenotypes in a polymorphic population evolves over time. For instance, under the replicator equation (Taylor and Jonker 1978) of the standard polymorphic population model (i.e. each phenotype existing in the population is a pure strategy), the evolutionary outcome is characterized as a locally asymptotically stable rest point of this dynamical system.

For classical matrix games (i.e. matrix games without time constraints), these two concepts are connected by one part of the folk theorem of evolutionary game theory (Hofbauer and Sigmund, 1998; Cressman, 2003; Broom and Rychtar, 2013): An ESS is a locally asymptotically stable rest point

of the replicator equation. The fundamental question of this paper is then: What is the connection between ESSs and stable rest points of the standard replicator equation in the class of matrix games under time constraints?

To address this question, we first show in Section 3 (Proposition 3.1) that playing the game against a polymorphic population is equivalent to interacting with a specific monomorphism whose phenotype is the average strategy of the active individuals in the polymorphic population. Moreover, for the standard polymorphic model of Section 4, there is a unique distribution of individual phenotypes in the population that corresponds to this monomorphism.[1]

In the special case that all individuals in the population are playing the same pure strategy, the monomorphism is this pure strategy. It is then straightforward to show (Theorem 4.1) that a strict Nash equilibrium (NE) according to Definition 2.2 is an ESS that is locally asymptotically stable under the replicator equation, thereby generalizing another part of the folk theorem of evolutionary game theory to matrix games with time constraints.

Our main result (Theorem 4.2) states that, when there are two pure strategies, the distribution in the standard polymorphic model is locally asymptotically stable under the replicator equation if and only if its corresponding monomorphism is an ESS. That is, the two solution concepts of evolutionary game theory are equivalent for two-strategy games with time constraints. The lengthy proof of Theorem 4.2 relies heavily on special techniques for two-strategy games that do not generalize to higher dimensional strategy spaces. In fact, although an ESS still corresponds to a polymorphic rest point of the replicator equation (see Lemma 3.2), we conjecture that counterexamples to stability of this polymorphic distribution already appear in three-strategy games.

2 Matrix games under time constraints and the monomorphic model

In a matrix game, there are n pure strategies and an individual's phenotype is given by a mixed strategy (i.e. a probability distribution $\mathbf{p} =$

---

[1]In classical matrix games, this monomorphism is the same as the mixed strategy given by the polymorphic population. This is no longer true in general when the effect of time constraints is considered and there are two or more pure strategies in use by the polymorphic population.

$(p_1, \ldots, p_n) \in S_n = \{\mathbf{p} \in \mathbb{R}^n : 0 \le p_i \le 1, \sum p_i = 1\}$ on these pure strategies) whereby it uses pure strategy i with probability p_i at a given time. In our time-constrained model, individuals can either be active or inactive. An active individual meets other individuals at random at a fixed rate. It plays a two-player (symmetric) game when it encounters another active individual, receiving an intake that depends on its strategy and that of its opponent. After encountering another active individual, the active individual becomes inactive for a certain amount of time that also depends on its strategy and that of its opponent. This positive amount of time may include the time it takes to play the game (i.e. interaction time), a waiting or recovery time after the interaction before it is ready to play another game, or a handling time if the interaction models competition over a resource (e.g. a foraging game).

If, at a given time, the focal active individual using the i-th pure strategy meets an active opponent who is using the j-th pure strategy, then the focal individual's intake is a_{ij} and the focal individual cannot play the next game during an average time duration $\tau_{ij} > 0$. Hence our time-constrained matrix game is characterized by two matrices, the intake matrix $A = (a_{ij})_{n \times n}$, and the time constraint matrix $T = (\tau_{ij})_{n \times n}$. For individual fitness in this game, we follow Garay et al. (2017) who assume a continuous time Markov model is used where a focal individual's time between encounters when active and the amount of time it is inactive are independent and exponentially distributed with prescribed mean.[2] For a large population in this situation, they show that the fitness of the focal individual is given by the quotient

$$W = \frac{E(A)}{E(T)}, \tag{1}$$

where $E(A)$ and $E(T)$ are the average intake and average time, evaluated at the stationary distribution of the Markov process, for the focal individual's phenotype during one cycle of it being active and inactive (i.e. one activity cycle).[3]

---

[2]That is, we assume each activity (i.e. active or inactive) can happen during a random time duration, which is exponentially distributed. This is quite different from the situation where the time duration of an action is part of the strategy of the player, for instance, the war of attrition (e.g. Eriksson et al. 2004), and dispersal-foraging game (Garay et al. 2015a).

[3]Observe that the Holling functional response are defined in a similar way (Holling

For the monomorphic model, we follow the setup of Maynard Smith (1982). That is, the resident population is monomorphic whereby every individual has resident phenotype $\mathbf{p}^*$. To be a stable evolutionary outcome, $\mathbf{p}^*$ must resist invasion by a small mutant subpopulation whose members all use a mutant phenotype $\mathbf{p}$ that is different from $\mathbf{p}^*$. It does this by having the higher fitness in the resident-mutant system whenever the proportion of mutants is small enough. Let ε and $(1-\varepsilon)$ be the frequencies of mutant and resident phenotypes in the resident-mutant system. Since the population is large and well mixed, a focal active individual (independent of its phenotype) meets an active (respectively, inactive) mutant with rate $\varepsilon\rho$ (respectively, $\varepsilon(1-\rho)$) and an active (respectively, inactive) resident with rate $(1-\varepsilon)\rho^*$ (respectively, $(1-\varepsilon)(1-\rho^*)$) where ρ (respectively, ρ^*) is the relative frequency of mutants (respectively, residents) who are active.[4] That is, the encounter distribution of the focal individual is $(\varepsilon\rho,\ \varepsilon(1-\rho),\ (1-\varepsilon)\rho^*,\ (1-\varepsilon)(1-\rho^*))$. Moreover, as shown by Garay et al. (2017), the stationary distribution of the Markov process for fixed ε satisfies the system of equations

$$\rho\ =\ \frac{1}{1 + \mathbf{p}T[(1-\varepsilon)\rho^*\mathbf{p}^* + \varepsilon\rho\mathbf{p}]}$$

$$\rho^*\ =\ \frac{1}{1 + \mathbf{p}^*T[(1-\varepsilon)\rho^*\mathbf{p}^* + \varepsilon\rho\mathbf{p}]}.$$

(2)

where, for instance, $\mathbf{p}T(1-\varepsilon)\rho^*\mathbf{p}^* = (1-\varepsilon)\rho^* \sum_{i,j=1}^{n} p_i \tau_{ij} p_j^*$ is the probability a focal active mutant meets an active resident times the expected amount of time the mutant is then inactive. They also show this system[5] has a unique solution (ρ, ρ^*) in $[0,1] \times [0,1]$ (see Lemma 6.1 in A.1).

1959, Garay and Móri 2010). Also, observe that we do not use the term "payoff" for matrix games under time constraints. Instead, we use "intake" for the entries a_{ij} and "fitness" for (1) to avoid ambiguity since both these concepts have been called payoff in other circumstances.

[4]This assumes, without loss of generality, that the time unit is chosen in such a way that the fixed rate an active individual meets an individual at random is 1.

[5]The numerator 1 on the right-hand sides of (2) is the length of the active part of an activity cycle and the denominator is the expected time of an activity cycle for the focal individual whose phenotype is $\mathbf{p}$ (first equation) or $\mathbf{p}^*$ (second equation). That is, the left-hand and right-hand sides are both equal to the proportion of active individuals in the mutant (first equation) and resident (second equation) population. Note that ρ and ρ^* depend on $\mathbf{p}^*, \mathbf{p}$ and ε. Therefore, if it is necessary to emphasize this dependence, we use the notations $\rho_{\mathbf{p}}(\mathbf{p}^*, \mathbf{p}, \varepsilon)$ and $\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{p}, \varepsilon)$, respectively, instead of ρ and ρ^*, respectively.

From (1), the individual fitness of a resident and a mutant is given by

$$\omega^* := \frac{\mathbf{p}^*A[(1-\varepsilon)\rho^*\mathbf{p}^* + \varepsilon\rho\mathbf{p}]}{1 + \mathbf{p}^*T[(1-\varepsilon)\rho^*\mathbf{p}^* + \varepsilon\rho\mathbf{p}]} \text{ and } \omega := \frac{\mathbf{p}A[(1-\varepsilon)\rho^*\mathbf{p}^* + \varepsilon\rho\mathbf{p}]}{1 + \mathbf{p}T[(1-\varepsilon)\rho^*\mathbf{p}^* + \varepsilon\rho\mathbf{p}]},$$

respectively, evaluated at the stationary distribution. To emphasize that ω and ω^* depend on $\mathbf{p}^*, \mathbf{p}$ and ε, we use the notations $\omega_{\mathbf{p}}(\mathbf{p}^*, \mathbf{p}, \varepsilon)$ and $\omega_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{p}, \varepsilon)$, respectively, if it is necessary. From (2), the resident phenotype will have higher fitness than the mutant (i.e. $\omega^* > \omega$) at a given ε if and only if

$$\rho^*\mathbf{p}^*A[(1-\varepsilon)\rho^*\mathbf{p}^* + \varepsilon\rho\mathbf{p}] > \rho\mathbf{p}A[(1-\varepsilon)\rho^*\mathbf{p}^* + \varepsilon\rho\mathbf{p}]. \tag{3}$$

This has the following biological interpretation: the active resident phenotype $\mathbf{p}^*$ has higher intake against the whole active population than the active mutant phenotype $\mathbf{p}$.

Other equivalent interpretations emerge by considering the fitness $\bar{\omega} = \bar{\omega}(\mathbf{p}^*, \mathbf{p}, \varepsilon) := (1-\varepsilon)\omega^* + \varepsilon\omega$ of a random individual chosen in the resident-mutant population. It is straightforward to show that (3) is equivalent to either of the inequalities $\omega^* > \bar{\omega}$ or $\bar{\omega} > \omega$. That is, the resident phenotype has higher than the average fitness of the whole population or, alternatively, the mutant phenotype has lower fitness than the average fitness of the whole population. As we will see, these different views will be important in the paper.

From the monomorphic approach of Maynard Smith (1982), we have the following definition (see also Garay et al. (2017)).

Definition 2.1 *A* $\mathbf{p}^* \in S_n$ *is an **evolutionarily stable strategy of the matrix game under time constraints** (ESS), if, for all* $\mathbf{p} \neq \mathbf{p}^*$, *there exists an* $\varepsilon_0 = \varepsilon_0(\mathbf{p})$ *such that (3) holds whenever* $0 < \varepsilon < \varepsilon_0$.

That is, $\mathbf{p}^*$ is an ESS if and only if it resists invasion by any mutant phenotype if the mutant is initially sufficiently rare in the resident-mutant system. The ESS concept is also equivalent to dynamic stability in the following sense. The replicator dynamics for the resident-mutant system has the form

$$\begin{aligned}\dot{\varepsilon} &= \varepsilon(\omega - \bar{\omega}) = \varepsilon(1-\varepsilon)(\omega - \omega^*) \\ &= \varepsilon(1-\varepsilon)\left(\rho\mathbf{p}A[(1-\varepsilon)\rho^*\mathbf{p}^* + \varepsilon\rho\mathbf{p}] - \rho^*\mathbf{p}^*A[(1-\varepsilon)\rho^*\mathbf{p}^* + \varepsilon\rho\mathbf{p}]\right).\end{aligned}$$

This replicator dynamics satisfies the Darwinian tenet that the frequency of the mutant phenotype decreases if its fitness ω is less than the mean fitness $\bar{\omega}$ of the resident mutant system. From Definition 2.1 and inequality (3), strategy $\mathbf{p}^* \in S_n$ is an ESS if and only if $\varepsilon = 0$ is a locally asymptotically stable equilibrium of this replicator dynamics for all $\mathbf{p} \in S_n$ with $\mathbf{p} \neq \mathbf{p}^*$.

As pointed out by Garay et al. (2017), the main obstacle to an analytical formula for an ESS in terms of the matrices A and T of the time-constrained game is that the stationary distribution of the population from (2), which depends in a complicated way on the time constraint matrix T as well as the strategies of mutant and resident phenotypes for fixed ε, must be substituted into (3). On the other hand, since the fitness functions on both sides of inequality (3) depend continuously on the frequencies of the phenotypes (see Corollary 6.3 in A.1) as $\varepsilon \to 0$, an ESS satisfies inequality (4) in the following definition of a Nash equilibrium (NE). That is, an ESS is a NE.

Definition 2.2 *A strategy $\mathbf{p}^* \in S_N$ is a **Nash equilibrium of the matrix game under time constraints** (NE), if, for every $\mathbf{p} \neq \mathbf{p}^*$,*

$$
\begin{aligned}
\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{p}, 0)\mathbf{p}^* A \mathbf{p}^* &= \frac{\mathbf{p}^* A \mathbf{p}^*}{1 + \mathbf{p}^* T \rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{p}, 0)\mathbf{p}^*} \\
&\geq \frac{\mathbf{p} A \mathbf{p}^*}{1 + \mathbf{p} T \rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{p}, 0)\mathbf{p}^*} \\
&= \rho_{\mathbf{p}}(\mathbf{p}^*, \mathbf{p}, 0)\mathbf{p} A \mathbf{p}^*.
\end{aligned}
\tag{4}
$$

If the previous inequality is strict (for every $\mathbf{p} \neq \mathbf{p}^*$) we say that $\mathbf{p}^*$ is a ***strict NE***.

Note that, by continuity of the fitness functions, a strict NE is automatically an ESS. However, a strict NE is necessarily a pure strategy in S_n (see Theorem 4.1 below) and so there are many ESSs that are not strict NE. To see this, take all entries in T to be equal to τ. Then $\mathbf{p}^* T[(1 - \varepsilon)\rho^* \mathbf{p}^* + \varepsilon\rho\mathbf{p}] = \tau[(1 - \varepsilon)\rho^* + \varepsilon\rho] = \mathbf{p} T[(1 - \varepsilon)\rho^* \mathbf{p}^* + \varepsilon\rho\mathbf{p}]$. Thus, from (2), $\rho^* = \rho = (-1 + \sqrt{1 + 4\tau})/(2\tau)$ since $\rho = 1/(1 + \tau\rho)$. Furthermore, inequality (3) is equivalent to $\mathbf{p}^* A[(1 - \varepsilon)\mathbf{p}^* + \varepsilon\mathbf{p}] > \mathbf{p} A[(1 - \varepsilon)\mathbf{p}^* + \varepsilon\mathbf{p}]$ and so an ESS according to Definition 2.1 is the same as the classical concept of ESS for a matrix game A with no time constraint. It is well-known (Hofbauer and Sigmund, 1998) that many such games have mixed strategy ESSs.

3 Monomorphic versus polymorphic populations

In the polymorphic population, there is a fixed finite set of phenotypes taken from the phenotype pool S_n, at least two of which have positive frequency. Let the phenotypes be denoted by $\mathbf{p}^*$ and $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m$.[6] For $i = 1, 2, ..., m$, let x_i be the proportion of phenotype $\mathbf{p}_i$ in the population and set $x = x_1 + ... + x_m$. Then the proportion of phenotype $\mathbf{p}^*$ is $1 - x \geq 0$. Let ϱ^* and ϱ_i for $i = 1, 2, ..., m$ denote the proportions of active individuals within the phenotype $\mathbf{p}^*$ and phenotype $\mathbf{p}_i$ populations, respectively.

We assume that, for a given $x_1, ..., x_m$, the polymorphic system is at its stationary distribution. Generalizing (2), this is the unique solution $(\varrho^*, \varrho_1, ..., \varrho_m)$ in the unit hypercube $[0, 1]^{m+1}$ (Garay et al., 2017) of the following system of equations.

$$\varrho^* = \frac{1}{1 + \mathbf{p}^* T[(1 - x)\varrho^* \mathbf{p}^* + \sum_{j=1}^{m} x_j \varrho_j \mathbf{p}_j]} \tag{5}$$

$$\varrho_i = \frac{1}{1 + \mathbf{p}_i T[(1 - x)\varrho^* \mathbf{p}^* + \sum_{j=1}^{m} x_j \varrho_j \mathbf{p}_j]} \quad 1 \leq i \leq m.$$

Furthermore, the average intakes per time unit (i.e. fitness given as in (1)) of phenotype $\mathbf{p}^*$ and phenotype $\mathbf{p}_i$, respectively, are

$$W^* = \varrho^* \mathbf{p}^* A \left[(1 - x)\varrho^* \mathbf{p}^* + \sum_{j=1}^{m} x_j \varrho_j \mathbf{p}_j \right] \tag{6}$$

$$W_i = \varrho_i \mathbf{p}_i A \left[(1 - x)\varrho^* \mathbf{p}^* + \sum_{j=1}^{m} x_j \varrho_j \mathbf{p}_j \right] \quad 1 \leq i \leq m.$$

It is important to note that $\varrho^*, \varrho_i, W^*$ and W_i depend on the frequency distribution

$$\mathbf{x} := (1 - x, x_1, \ldots, x_m) \in S_{m+1}.$$

Therefore the notations $\varrho^*(\mathbf{x}), \varrho_i(\mathbf{x}), W^*(\mathbf{x})$ and $W_i(\mathbf{x})$, respectively, are used if we want to emphasize this dependence.

---

[6]In the following, $\mathbf{p}^*$ is often thought of as the resident strategy or as an ESS according to Definition 2.1. However, since this does not need to be the case, it is better to regard $\mathbf{p}^*$ only as a phenotype that is distinguished from the others.

Our first main result (Proposition 3.1) is that, from the point of view of phenotype $\mathbf{p}^*$, the collection of other phenotypes $\mathbf{p}_i$ with proportions given through $\mathbf{x}$ in the large polymorphic population can be replaced with a single mixed phenotype as in the monomorphic model of Section 2. For this purpose, we define

$$\tilde{\varrho}(\mathbf{x}) = \frac{1}{x} \sum_{j=1}^{m} x_j \varrho_j(\mathbf{x}) \tag{7}$$

$$\tilde{\mathbf{h}}(\mathbf{x}) = \frac{\sum_{i=1}^{m} x_i \varrho_i(\mathbf{x}) \mathbf{p}_i}{\sum_{j=1}^{m} x_j \varrho_j(\mathbf{x})} = \frac{1}{x \tilde{\varrho}(\mathbf{x})} \sum_{i=1}^{m} x_i \varrho_i(\mathbf{x}) \mathbf{p}_i. \tag{8}$$

That is, $\tilde{\varrho}$ is the average frequency of active individuals among the phenotypes $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m$ and $\tilde{\mathbf{h}}(\mathbf{x})$ is the average strategy of these active individuals.[7] The result is

Proposition 3.1 *Consider a large polymorphic population with probability distribution* $\mathbf{x}$. *At the stationary distribution of this polymorphic system,* $\varrho^*(\mathbf{x})$ *and* $W^*(\mathbf{x})$ *in (5) and (6) respectively are given by the monomorphic model based on phenotypes* $\mathbf{p}^*$ *and* $\tilde{\mathbf{h}}(\mathbf{x})$ *with proportions* $1 - x$ *and* x, *respectively. That is,* $\rho_{\mathbf{p}^*}(\mathbf{p}^*, \tilde{\mathbf{h}}(\mathbf{x}), x) = \varrho^*(\mathbf{x})$ *and* $\omega_{\mathbf{p}^*}(\mathbf{p}^*, \tilde{\mathbf{h}}(\mathbf{x}), x) = W^*(\mathbf{x})$.

Furthermore, $\varrho^*(\mathbf{x})$ *and* $\tilde{\varrho}(\mathbf{x})$ *give the stationary distribution for the monomorphic model (in particular,* $\rho_{\tilde{\mathbf{h}}(\mathbf{x})}(\mathbf{p}^*, \tilde{\mathbf{h}}(\mathbf{x}), x) = \tilde{\varrho}(\mathbf{x})$) *and the fitness of* $\tilde{\mathbf{h}}(\mathbf{x})$ *is the average fitness* $\tilde{W}(\mathbf{x}) \equiv \frac{1}{x} \sum_{j=1}^{m} x_i W_i(\mathbf{x})$ *of the phenotypes* $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m$ *in the polymorphic model (i.e.* $\omega_{\tilde{\mathbf{h}}(\mathbf{x})}(\mathbf{p}^*, \tilde{\mathbf{h}}(\mathbf{x}), x) = \tilde{W}(\mathbf{x})$).

Proof. Recall the defining equations (5) of ϱ^* and ϱ_i, respectively:

$$\varrho^* = \frac{1}{1 + \mathbf{p}^* T[(1-x)\varrho^* \mathbf{p}^* + \sum_{j=1}^{m} x_j \varrho_j \mathbf{p}_j]} \tag{9}$$

$$\varrho_i = \frac{1}{1 + \mathbf{p}_i T[(1-x)\varrho^* \mathbf{p}^* + \sum_{j=1}^{m} x_j \varrho_j \mathbf{p}_j]} \qquad 1 \le i \le m. \tag{10}$$

Multiplying both sides of (10) by its denominator and then by x_i/x yields:

$$\frac{1}{x} \left[x_i \varrho_i + x_i \varrho_i \mathbf{p}_i T \left[(1-x)\varrho^* \mathbf{p}^* + \sum_{j=1}^{m} x_j \varrho_j \mathbf{p}_j \right] \right] = \frac{x_i}{x}. \tag{11}$$

---

[7]Note that, $\tilde{\varrho}(\mathbf{x})$ and $\tilde{\mathbf{h}}(\mathbf{x})$ are well-defined since at least one of the x_i is positive.

Recall the definition of $\tilde{\varrho}$ in (7) and note that, by (8), it is true that

$$\sum_{j=1}^{m} x_j \varrho_j \mathbf{p}_j = x\tilde{\varrho}\frac{1}{x\tilde{\varrho}}\sum_{j=1}^{m} x_j \varrho_j \mathbf{p}_j = x\tilde{\varrho}\tilde{\mathbf{h}}.$$

Therefore, if we sum equations (11) from $i = 1$ to $i = m$ we receive that

$$\tilde{\varrho} + \tilde{\varrho}\tilde{\mathbf{h}}T\left[(1-x)\varrho^*\mathbf{p}^* + x\tilde{\varrho}\tilde{\mathbf{h}}\right] = 1 \tag{12}$$

which, divided by $1 + \tilde{\mathbf{h}}T[(1-x)\varrho^*\mathbf{p}^* + x\tilde{\varrho}\tilde{\mathbf{h}}]$, is equivalent to

$$\tilde{\varrho} = \frac{1}{1 + \tilde{\mathbf{h}}T\left[(1-x)\varrho^*\mathbf{p}^* + x\tilde{\varrho}\tilde{\mathbf{h}}\right]}, \tag{13}$$

and so $\tilde{\varrho}(\mathbf{x}) = \rho_{\tilde{\mathbf{h}}(\mathbf{x})}(\mathbf{p}^*, \tilde{\mathbf{h}}(\mathbf{x}), x)$.

Similarly, write (6) in the following form:

$$\begin{aligned}
W^* &= \varrho^*\mathbf{p}^*A[(1-x)\varrho^*\mathbf{p}^* + x\tilde{\varrho}\tilde{\mathbf{h}}] \\
W_i &= \varrho_i\mathbf{p}_iA[(1-x)\varrho^*\mathbf{p}^* + x\tilde{\varrho}\tilde{\mathbf{h}}] \quad 1 \leq i \leq m.
\end{aligned}$$

Mimic the previous calculation for the "ϱ"-s from (11) to (12) to get

$$\tilde{W} = \tilde{\varrho}\tilde{\mathbf{h}}A[(1-x)\varrho^*\mathbf{p}^* + x\tilde{\varrho}\tilde{\mathbf{h}}]$$

which completes the proof. ■

We will use Proposition 3.1 to investigate how an ESS $\mathbf{p}^*$ of the monomorphic model is related to equilibrium and stability in two different polymorphic models. In this section, we focus on the case where $\mathbf{p}^*$ is the resident phenotype and the frequencies of the other phenotypes $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m$ (now called mutants) are small.[8] This models the situation where mutations are not rare events since there can be more than one type of mutant in the system at a given time. Maynard Smith (1982, Appendix D) has already called attention to the possibility in the classical matrix games that a resident strategy, which cannot be invaded by any single mutant using a pure strategy, can

---

[8]Section 4 considers the pure-strategy polymorphic model where the only phenotypes present in the population use one of the n pure strategies.

sometimes be invaded when two mutant pure strategies are simultaneously present. This continues to be true under arbitrary time constraint as we shall show shortly.

To consider this fact, suppose that the resident phenotype is in the convex hull of $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m$. That is,

$$\mathbf{p}^* = \sum_{i=1}^{m} \alpha_i \mathbf{p}_i, \text{ where all } \alpha_i \geq 0 \text{ and } \sum_{i=1}^{m} \alpha_i = 1.$$

Define $\varrho^*(0)$ and $\varrho_i(0)$ as the unique positive solution of (5) when $x = 0$. That is,

$$\varrho^*(0) = \frac{1}{1 + \mathbf{p}^* T \varrho^*(0) \mathbf{p}^*}$$

$$\varrho_i(0) = \frac{1}{1 + \mathbf{p}_i T \varrho^*(0) \mathbf{p}^*} \qquad 1 \leq i \leq m.$$

Now take $\hat{x}_i(x) = x \alpha_i \varrho^*(0) / \varrho_i(0)$ for an arbitrary $0 \leq x \leq 1$. Then

$$\begin{aligned}
\sum_{i=1}^{m} \hat{x}_i(x) &= x \sum_{i=1}^{m} \alpha_i \frac{\varrho^*(0)}{\varrho_i(0)} = x \sum_{i=1}^{m} \alpha_i \varrho^*(0) \left[1 + \mathbf{p}_i T \varrho^*(0) \mathbf{p}^* \right] \\
&= x \varrho^*(0) \left(1 + \mathbf{p}^* T \varrho^*(0) \mathbf{p}^* \right) = x
\end{aligned}$$

and so $\hat{\mathbf{x}}(x) \equiv (1 - x, \hat{x}_1(x), \ldots, \hat{x}_m(x)) \in S_{m+1}$. Since $T[(1 - x)\varrho^*(0)\mathbf{p}^* + \sum_{j=1}^{m} \hat{x}_j(x)\varrho_j(0)\mathbf{p}_j] = T[(1 - x)\varrho^*(0)\mathbf{p}^* + \sum_{j=1}^{m} x\alpha_j \frac{\varrho^*(0)}{\varrho_j(0)}\varrho_j(0)\mathbf{p}_j] = T\varrho^*(0)\mathbf{p}^*$, the system (5) for the distribution $\hat{\mathbf{x}}(x)$ satisfies

$$\varrho^*(0) = \frac{1}{1 + \mathbf{p}^* T[(1 - x)\varrho^*(0)\mathbf{p}^* + \sum_{j=1}^{m} \hat{x}_j(x)\varrho_j(0)\mathbf{p}_j]} = \frac{1}{1 + \mathbf{p}^* T \varrho^*(0) \mathbf{p}^*}$$

$$\varrho_i(0) = \frac{1}{1 + \mathbf{p}_i T[(1 - x)\varrho^*(0)\mathbf{p}^* + \sum_{j=1}^{m} \hat{x}_j(x)\varrho_j(0)\mathbf{p}_j]} = \frac{1}{1 + \mathbf{p}_i T \varrho^*(0) \mathbf{p}^*} \qquad 1 \leq i \leq m.$$

That is, the unique solution to (5) for $\hat{\mathbf{x}}(x)$ is $(\varrho^*(0), \varrho_1(0), ..., \varrho_m(0))$. In particular, $\varrho^*(\hat{\mathbf{x}}(x)) = \varrho^*(0)$ and $\varrho_j(\hat{\mathbf{x}}(x)) = \varrho_j(0)$. From this, it follows that

$$\tilde{\varrho}(\hat{\mathbf{x}}(x)) = \frac{1}{x} \sum_{j=1}^{m} \hat{x}_j(x)\varrho_j(\hat{\mathbf{x}}(x)) = \frac{1}{x} x \sum_{j=1}^{m} \alpha_j \frac{\varrho^*(0)}{\varrho_j(0)}\varrho_j(0) = \varrho^*(0)$$

$$\tilde{\mathbf{h}}(\mathbf{x}(x)) = \frac{1}{x\tilde{\varrho}(\hat{\mathbf{x}}(x))} \sum_{i=1}^{m} \hat{x}_i(x)\varrho_i(\hat{\mathbf{x}}(x))\mathbf{p}_i = \frac{x}{x\varrho^*(0)} \sum_{i=1}^{m} \alpha_i \frac{\varrho^*(0)}{\varrho_i(0)}\varrho_i(0)\mathbf{p}_i = \mathbf{p}^*.$$

13

Thus, by Proposition 3.1

$$W^*(\hat{\mathbf{x}}(x)) = \omega_{\mathbf{p}^*}(\mathbf{p}^*, \tilde{\mathbf{h}}(\hat{\mathbf{x}}(x)), x) = \omega_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{p}^*, x)$$
$$= \omega_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{p}^*, 0) = W^*(\hat{\mathbf{x}}(0))$$

and

$$\tilde{W}(\hat{\mathbf{x}}(x)) = \omega_{\tilde{\mathbf{h}}(\hat{\mathbf{x}}(x))}(\mathbf{p}^*, \tilde{\mathbf{h}}(\tilde{\mathbf{x}}(x)), x) = \omega_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{p}^*, x)$$
$$= \omega_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{p}^*, 0) = W^*(\hat{\mathbf{x}}(0)),$$

respectively. In summary, if $\mathbf{p}^*$ is a convex combination of $\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_m$, then there is always a state $\hat{\mathbf{x}}(x) \in S_{m+1}$ for any $x \in [0, 1]$ in the polymorphic model with $W^*(\hat{\mathbf{x}}(0)) = W^*(\hat{\mathbf{x}}(x)) = \tilde{W}(\hat{\mathbf{x}}(x))$. In particular, for this polymorphic model, it is always possible that a combination of the mutant phenotypes has the same average fitness $\tilde{W}$ as the fitness W^* of the resident strategy $\mathbf{p}^*$ no matter how small the total frequency of mutant phenotypes is.[9]

In this sense, a monomorphic ESS $\mathbf{p}^*$ can be invaded in the polymorphic model. For this reason, we turn our attention to the standard pure-strategy polymorphic model in the following section. Before doing so, it is important to mention that a NE $\mathbf{p}^*$ does correspond to an equilibrium of the polymorphic model by the following result.

Lemma 3.2 *Suppose $\mathbf{p}^*$ is a NE according to Definition 2.2. Then $W^*(\hat{\mathbf{x}}(0))$ $\geq W_i(\hat{\mathbf{x}}(0))$ for all $i = 1, ..., m$. Moreover, $W_i(\hat{\mathbf{x}}(x)) = W^*(\hat{\mathbf{x}}(0))$ for every $x \in [0, 1]$ whenever $\hat{x}_i(x) = x\alpha_i \varrho^*(0)/\varrho_i(0) > 0$. That is, $\hat{\mathbf{x}}(x)$ is an equilibrium of the polymorphic model.*

Proof. From Definition 2.2, $W^*(\hat{\mathbf{x}}(0)) = \rho^*(0)\mathbf{p}^* A\mathbf{p}^*\rho^*(0) \geq \rho_i(0)\mathbf{p}_i A\mathbf{p}^*\rho^*(0) = W_i(\hat{\mathbf{x}}(0))$ for all $i = 1, ..., m$. On the other hand, we have just seen that $W^*(\hat{\mathbf{x}}(0)) = W^*(\hat{\mathbf{x}}(x)) = \tilde{W}(\hat{\mathbf{x}}(x)) = \sum_{i=1}^m \hat{x}_i(x)W_i(\hat{\mathbf{x}}(x))$ for every $x \in [0, 1]$. This is possible if and only if $W_i(\hat{\mathbf{x}}(x)) = W^*(\hat{\mathbf{x}}(0))$ for every $x \in [0, 1]$ whenever $\hat{x}_i(x) > 0$. ∎

---

[9]This generalizes the classic result (Cressman 1992; Hofbauer and Sigmund 1998) that, in a mixed-strategy model, the average fitness of mutant phenotypes equals the fitness of their convex combination.

4 Evolutionary and Dynamic Stability in the Pure-Strategy Model

In the standard polymorphic population, there are $m = n$ phenotypes and each individual uses one of the pure strategies $\mathbf{e}_i \in S_n$ where $\mathbf{e}_i$ denotes the n-dimensional vector the i-th coordinate of which is 1 and the others are 0. Since we do not distinguish a strategy[10] in the polymorphic population in the present reasoning and we need the average frequency of active individuals and the average strategy of active individuals, we define $\bar{\varrho}$ and $\bar{\mathbf{h}}$ copying the definition of $\tilde{\varrho}$ and $\tilde{\mathbf{h}}$ in (7) and (8) as follows:

$$\bar{\varrho}(\mathbf{x}) = \sum_{i=1}^{n} x_i \varrho_i(\mathbf{x}) \tag{14}$$

and

$$\frac{1}{\bar{\varrho}(\mathbf{x})} \sum_{i=1}^{n} x_i \varrho_i(\mathbf{x}) \mathbf{e}_i \tag{15}$$

respectively, where $\mathbf{x} \in [0,1]^n$ with $\sum x_i = 1$. In other words, $\bar{\varrho}(\mathbf{x}) = \tilde{\varrho}(0, \mathbf{x}) = \tilde{\varrho}(0, x_1, \ldots, x_n)$ and $\bar{\mathbf{h}}(\mathbf{x}) = \tilde{\mathbf{h}}(0, \mathbf{x})$. Then there is a unique $\mathbf{x} \in S_n$ with $\bar{\mathbf{h}}(\mathbf{x}) = \mathbf{p}^*$ (see Corollary 6.3 (iii) and (iv) in A.1).This is given by the frequency distribution

$$x_i = \frac{\rho^*}{\rho_i} p_i^* \tag{16}$$

where $\rho^* = \rho_{\mathbf{p}^*}$ is the unique solution in $[0,1]$ to the equation

$$\rho^* = \frac{1}{1 + \mathbf{p}^* T \rho^* \mathbf{p}^*} \tag{17}$$

and $\rho_i = \rho_i(\mathbf{p}^*)$ is the expression

$$\frac{1}{1 + \mathbf{e}_i T \rho^* \mathbf{p}^*} \qquad 1 \leq i \leq n. \tag{18}$$

Now consider the standard replicator equation

$$\dot{x}_i = x_i [W_i(\mathbf{x}) - \bar{W}(\mathbf{x})], \tag{19}$$

---

[10]see footnote 6 on page 10

for the pure-strategy model where, as before, $W_i(\mathbf{x})$, $i \in \{1, ..., n\}$, is the fitness of the i-th phenotype and $\bar{W}(\mathbf{x}) = \sum x_i W_i(\mathbf{x})$ is the average fitness of the whole polymorphic population. From Lemma 3.2 (apply it with $x = 1$), if $\mathbf{p}^*$ is an ESS of the monomorphic model with $\bar{\mathbf{h}}(\mathbf{x}) = \mathbf{p}^*$, then $W_i(\mathbf{x}) = \bar{W}(\mathbf{x})$ whenever $x_i > 0$ and so $\mathbf{x}$ is a rest point of the replicator equation. The question of most interest now is whether such an $\mathbf{x}$ is a stable equilibrium of the replicator equation. Our first result is that this is true if $\mathbf{p}^*$ is a strict NE (and so an ESS).[11]

Theorem 4.1 *A strict NE must be a pure strategy. Without loss of generality, suppose that $\mathbf{e}_1$ is a strict NE. Then the state $\mathbf{x}^* = (1, 0, ..., 0) \in S_{1+m}$ is a locally asymptotically stable rest point of the replicator equation (19).*

Proof. Assume that $\mathbf{p}^*$ is a strict NE. Then, according to Lemma 3.2, $\omega_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{e}_i, 0) = W^*(\hat{\mathbf{x}}(0)) = W_i(\hat{\mathbf{x}}(0)) = \omega_{\mathbf{e}_i}(\mathbf{p}^*, \mathbf{e}_i, 0)$ for every i with $p_i^* \neq 0$. (Note that $\hat{\mathbf{x}}(0) = \mathbf{x}^*$.) Since $\mathbf{p}^*$ is a strict NE, this is possible if and only if $\mathbf{p}^*$ is a pure strategy. It can be assumed without loss of generality that $\mathbf{p}^* = \mathbf{e}_1$.

We prove that if we consider the replicator dynamics for a population consisting of individuals using one of the (pairwise distinct) strategies $\mathbf{e}_1, \mathbf{p}_1, \ldots, \mathbf{p}_m$ then $\mathbf{x}^* = (1, 0, \ldots, 0) \in S_{1+m}$ is a locally asymptotically stable rest point. We are looking for a $\delta_i > 0$ such that

$$\frac{\mathbf{e}_1 A(1-x)\varrho^* \mathbf{e}_1 + \mathbf{e}_1 A\left[\sum_{j=1}^m x_j \varrho_j \mathbf{p}_j\right]}{1 + \mathbf{e}_1 T(1-x)\varrho^* \mathbf{e}_1 + \mathbf{e}_1 T\left[\sum_{j=1}^m x_j \varrho_j \mathbf{p}_j\right]}$$
$$- \frac{\mathbf{p}_i A(1-x)\varrho^* \mathbf{e}_1 + \mathbf{p}_i A\left[\sum_{j=1}^m x_j \varrho_j \mathbf{p}_j\right]}{1 + \mathbf{p}_i T(1-x)\varrho^* \mathbf{e}_1 + \mathbf{p}_i T\left[\sum_{j=1}^m x_j \varrho_j \mathbf{p}_j\right]} > 0 \quad (20)$$

hold whenever $0 < ||\mathbf{x} - \mathbf{x}^*|| \leq \delta_i$. Since $\mathbf{e}_1$ is a strict Nash equilibrium we have

$$\frac{\mathbf{e}_1 A\varrho^*(\mathbf{x}^*)\mathbf{e}_1}{1 + \mathbf{e}_1 T\varrho^*(\mathbf{x}^*)\mathbf{e}_1} - \frac{\mathbf{p}_i A\varrho^*(\mathbf{x}^*)\mathbf{e}_1}{1 + \mathbf{p}_i T\varrho^*(\mathbf{x}^*)\mathbf{e}_1} > 0. \quad (21)$$

The left-hand side of (20) is continuous in $\mathbf{x}$ (see Corollary 6.3) and, at $\mathbf{x} = \mathbf{x}^*$, it is just equal to the left-hand side of (21). This continuity implies the existence of δ_i.

---

[11]The proof of Theorem 4.1 shows that this result generalizes to the polymorphic model of the previous section in that stability of a strict NE $\mathbf{p}^*$ continues to hold for the replicator equation extended to sets of mixed strategy phenotypes that include $\mathbf{p}^*$.

Now, let $\delta := \min(\delta_1, \ldots, \delta_m)$. Then (20) holds for every $1 \leq i \leq m$ whenever $0 < ||\mathbf{x} - \mathbf{x}^*|| \leq \delta$. Since $\bar{W}(\mathbf{x}) = (1-x)W^*(\mathbf{x}) + \sum_i x_i W_i(\mathbf{x})$, one can immediately conclude that $W^*(\mathbf{x}) > \bar{W}(\mathbf{x})$ whenever $0 < ||\mathbf{x} - \mathbf{x}^*|| \leq \delta$. ∎

Theorem 4.1 is a well-known result for evolutionary games without time constraints, forming one part of the folk theorem of evolutionary game theory (Hofbauer and Sigmund, 1998; Cressman, 2003). Another part of the folk theorem states that an interior ESS (i.e. all components p_i^* of $\mathbf{p}^*$ are positive) is globally asymptotically stable under the replicator equation. This is no longer true for time constrained games as the following two-strategy example shows.

Example 1. Consider the two-strategy game with time constraint matrix and intake matrix given by

$$T := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad A := \begin{pmatrix} 1 & 1 \\ c & d \end{pmatrix},$$

respectively, where c and d are defined below. For the pure-strategy model, $\varrho_i(x)$ for $i = 1, 2$ is the solution of the equation

$$\varrho_i = \frac{1}{1 + \mathbf{e}_i T[x \varrho_1 \mathbf{e}_1 + (1-x) \varrho_2 \mathbf{e}_2]}.$$

That is,

$$\varrho_1(x) = \frac{-1 + \sqrt{1 + 4x}}{2x}, \quad \varrho_2(x) = \frac{-1 + \sqrt{5 - 4x}}{2(1-x)}.$$

A state $(x, 1 - x)$ is an interior rest point of the two-dimensional replicator equation if and only if $W_1(x) = W_2(x)$; that is

$$\varrho_1(x)\mathbf{e}_1 A[x \varrho_1(x)\mathbf{e}_1 + (1-x)\varrho_2(x)\mathbf{e}_2] = \varrho_2(x)\mathbf{e}_2 A[x \varrho_1(x)\mathbf{e}_1 + (1-x)\varrho_2(x)\mathbf{e}_2].$$

Thus, $u \neq v$ in $(0, 1)$ are both rest points if

$$u\varrho_1(u)^2 + (1-u)\varrho_1(u)\varrho_2(u) = u\varrho_1(u)\varrho_2(u)c + (1-u)\varrho_2(u)^2 d$$
$$v\varrho_1(v)^2 + (1-v)\varrho_1(v)\varrho_2(v) = v\varrho_1(v)\varrho_2(v)c + (1-v)\varrho_2(v)^2 d$$

17

which gives us two equations with unknowns c, d. For instance, if $u = 1/4$ and $v = 1/2$, one can straightforwardly solve this system to find that $c = \sqrt{2} - 1$ while $d = 3 - \sqrt{2}$. Moreover, for these values of c, d,

$$\frac{d}{dx}[W_1(x) - W_2(x)]\big|_{x=1/4} = \frac{7\sqrt{2} - 10}{3} < 0$$

while

$$\frac{d}{dx}[W_1(x) - W_2(x)]\big|_{x=1/2} = \frac{2(\sqrt{3} - 1)(6 - 2\sqrt{3} - \sqrt{6})}{3} > 0.$$

Since the replicator equation for two-strategy games is the one-dimensional dynamics $\dot{x}_1 = x_1(1 - x_1)[W_1(\mathbf{x}) - W_2(\mathbf{x})]$, the state $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2) = (1/4, 3/4)$ is locally asymptotically stable but not globally because there is another interior rest point.

Note that, by Theorem 4.2 below, strategy

$$\mathbf{p}^* := \frac{1}{4}\frac{\varrho_1(1/4)}{\bar{\varrho}(1/4)}\mathbf{e}_1 + \frac{3}{4}\frac{\varrho_2(1/4)}{\bar{\varrho}(1/4)}\mathbf{e}_2 = \left(1 - \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$$

is an ESS of this two-strategy time constrained game. Furthermore, it is straightforward to show that $\mathbf{e}_1$ is a strict NE (and so an ESS) while $\mathbf{e}_2$ is not. This contrasts with the evolutionary game without time constraints and payoff matrix A where $\mathbf{e}_1$ and $\mathbf{e}_2$ are both strict NE and the only interior NE $\left(\frac{1}{2}, \frac{1}{2}\right)$ is not an ESS.

By the following result, local asymptotic stability under the replicator equation does correspond to an ESS for two-strategy games.

Theorem 4.2 *Given* $\mathbf{p}^* \in S_2$, *the unique solution* $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2) \in S_2$ *of* $\bar{\mathbf{h}}(\mathbf{x}) = \mathbf{p}^{*}$[12] *is locally asymptotically stable under the replicator equation (19) if and only if* $\mathbf{p}^*$ *is an ESS of the monomorphic model.*

Proof. By Lemma 6.5 in A.2, given $\mathbf{p}^* \in S_2$, there is an $\varepsilon_0 > 0$ such that

$$\hat{x}_1 W_1(\mathbf{x}) + \hat{x}_2 W_2(\mathbf{x}) > \bar{W}(\mathbf{x}) \tag{22}$$

---

[12] From (16), $\hat{x}_1 = p_1^* \frac{\rho_{\mathbf{p}^*}}{\rho_1(\mathbf{p}^*)}$ and $\hat{x}_2 = p_2^* \frac{\rho_{\mathbf{p}^*}}{\rho_2(\mathbf{p}^*)}$.

for all $\mathbf{x} = (x_1, x_2) \in S_2$ with $0 < |x_1 - \hat{x}_1| < \varepsilon_0$ if and only if $\mathbf{p}^*$ is an ESS. Since $\bar{W}(\mathbf{x}) = x_1 W_1(\mathbf{x}) + x_2 W_2(\mathbf{x})$, $\hat{x}_2 = 1 - \hat{x}_1$ and $x_2 = 1 - x_1$, this inequality is equivalent to

$$(\hat{x}_1 - x_1)[W_1(\mathbf{x}) - W_2(\mathbf{x})] > 0$$

which holds if and only if, with $\mathbf{x} \in S_2$, $W_1(\mathbf{x}) - W_2(\mathbf{x}) > 0$ if $x_1 < \hat{x}_1$ and $W_1(\mathbf{x}) - W_2(\mathbf{x}) < 0$ if $x_1 > \hat{x}_1$. From the replicator equation $\dot{x}_1 = x_1(1 - x_1)[W_1(\mathbf{x}) - W_2(\mathbf{x})]$, the stability of a rest point $\hat{\mathbf{x}}$ is determined by the sign of the expression $W_1(\mathbf{x}) - W_2(\mathbf{x})$: $\hat{\mathbf{x}}$ is asymptotically stable if and only if $W_1(\mathbf{x}) - W_2(\mathbf{x}) > 0$ if $x_1 < \hat{x}_1$ and $W_1(\mathbf{x}) - W_2(\mathbf{x}) < 0$ if $x_1 > \hat{x}_1$ whenever $x_1 \in [0, 1]$ is close enough to $\hat{x}_1$. ∎

Remark. The proof of Theorem 4.2 uses the same method as Hofbauer and Sigmund (1998) who showed an interior ESS is globally asymptotically stable under the replicator equation for n-strategy matrix games without time constraint. The key to this proof is inequality (22), which for n-strategy games where $\hat{\mathbf{x}} \cdot \mathbf{W}(\mathbf{x}) := \sum_{i=1}^{n} \hat{x}_i W_i(\mathbf{x})$ becomes

$$\hat{\mathbf{x}} \cdot \mathbf{W}(\mathbf{x}) > \mathbf{x} \cdot \mathbf{W}(\mathbf{x}) \tag{23}$$

for all $\mathbf{x} \in S_n$ sufficiently close (but not equal) to $\hat{\mathbf{x}}$. The biological interpretation of inequality (23) is that, for all polymorphic population states near $\hat{\mathbf{x}}$, the average fitness of phenotype $\hat{\mathbf{x}}$ is higher than that of the whole population.

In Hofbauer and Sigmund (1998) (see especially their Section 6.5 on population games and equation (6.21)), such an $\hat{\mathbf{x}} \in S_n$ is defined as a local ESS (of the polymorphic pure-strategy model). To avoid confusion with our Definition 2.1 of an ESS for the monomorphic model, we will call an $\hat{\mathbf{x}}$ that satisfies (23) a polymorphic stable state (PSS) instead. Theorem 4.2 then states that, for two-strategy games, $\mathbf{p}^*$ is an ESS if and only if $\hat{\mathbf{x}}$ is a PSS (see Corollary 4.3 below). It is an open problem whether this equivalence extends to n-strategy matrix games with time constraint. In fact, we conjecture this equivalence is not true when $n > 2$ but have no counterexample at this time. On the other hand, the proof of Theorem 4.2 generalizes to show that a PSS is always asymptotically stable under the replicator equation for any n.

Corollary 4.3 *Let* $\mathbf{p}^* = (p_1^*, p_2^*) \in S_2$ *and let*

$$\hat{x}_i := p_i^* \frac{\rho^*}{\rho_i(\mathbf{p}^*)} \qquad i = 1, 2$$

where ρ^* and $\rho_i(\mathbf{p}^*)$ are defined in (17) and (18). Then the following three conditions are equivalent:

(i) $\mathbf{p}^*$ *is an ESS;*

(ii) $(\hat{x}_1, \hat{x}_2)$ *is a PSS;*

(iii) $(\hat{x}_1, \hat{x}_2)$ *is a locally asymptotically stable rest point of the replicator equation.*

5   Discussion

We have generalized two parts of the folk theorem of evolutionary game theory to the class of matrix games under time constraints. Specifically, Theorems 4.1 and 4.2 respectively show that a strict NE is locally asymptotically stable under the replicator equation and that, for two-strategy games, strategy $\hat{\mathbf{x}}$ is locally asymptotically stable under the replicator equation if and only if its corresponding monomorphism $\mathbf{h}(\hat{\mathbf{x}})$ is an ESS according to Definition 2.1.

Given the prominence of two-strategy games in applications of evolutionary game theory, these results mean that evolutionary outcomes of ecological models that incorporate the dynamic effects of different individual behaviors may continue to be predicted through static game-theoretic reasoning. In this regard, we emphasize that the replicator equation is an ecological model since the question it addresses is whether the long time existence of different strategies (i.e. different ecotypes) is possible. Moreover, the PSS defining condition (23) (which corresponds to ESS for two-strategy games by Corollary 4.3) is based on the ecological intuition that a state $\hat{\mathbf{x}}$ is stable if its fitness in a slightly perturbed ecological state is always higher than the average fitness in this perturbed state.[13]

---

[13]Another connection between ecology and our evolutionary dynamics is the fact that the replicator equation with n pure strategies in classical matrix games is equivalent to a Lotka-Volterra system with $n - 1$ species (Hofbauer & Sigmund 1998).

The effect of time constraints on evolutionary outcomes has been studied elsewhere in the literature. Besides our recent article (Garay et al 2017) that provides the foundation of the current investigation, Krivan and Cressman (2017) analyze general two-strategy matrix games when individuals are always paired, as in classical matrix games, but their interaction times depend on the strategies used in the pair. They were able to give an explicit formula for the stationary distribution for the standard polymorphic model in this case, in contrast to the implicit form we are forced to use in (2). On the other hand, our model extends theirs to include other activities beyond pair interactions that are essential to include for more realistic models of ecological systems.

References

[1] Broom, M, Luther, RM, Ruxton, GD and Rychtar, J 2008. A game-theoretic model of kleptoparasitic behavior in polymorphic populations. J Theor Biol 255, 81-91.

[2] Broom, M and Rychtar, J 2013. Game-Theoretical Models in Biology. Chapman & Hall/CRC Mathematical and Computational Biology

[3] Charnov, EL 1976. Optimal foraging: Attack strategy of a mantid. Am Nat 110, 141-151.

[4] Cressman R 1992. The Stability Concept of Evolutionary Game Theory (A Dynamic Approach). Lecture Notes in Biomathematics, vol 94, Springer-Verlag, Berlin

[5] Cressman R 2003. Evolutionary Dynamics and Extensive Form Games. MIT Press, Cambridge MASS.

[6] Eriksson, A, Lindgren, K and Lundh, T 2004. War of attrition with implicit time cost. J Theor Biol 230, 319-332.

[7] Garay J and Móri TF 2010. When is the opportunism remunerative? Community Ecol 11, 160-170.

[8] Garay J, Varga Z, Cabello T and Gámez M 2012. Optimal nutrient foraging strategy of an omnivore: Liebig's law determining numerical response. J Theor Biol 310, 31-42.

[9] Garay J, Cressman R, Xu F, Varga Z and Cabello T 2015a. Optimal Forager Against Ideal Free Distributed Prey. Am Nat 186, 111-122.

[10] Garay J, Varga Z, Gámez M and Cabello T 2015b. Functional response and population dynamics for fighting predator, based on activity distribution. J Theor Biol 368, 74-82.

[11] Garay J, Csiszár V and Móri, TF 2017. Evolutionary stability for matrix games under time constraints. J Theor Biol 415, 1-12.

[12] Hofbauer J and Sigmund K 1998. Evolutionary Games and Population Dynamics. Cambridge University Press

[13] Holling CS 1959. The components of predation as revealed by a study of small mammal predation of the European pine sawfly. The Canadian Entomologist 9, 293-320.

[14] Krivan V and Cressman R 2017. Interaction times change evolutionary outcomes: Two-player matrix games. J Theor Biol 416, 199-207.

[15] Maynard Smith J 1982. Evolution and the Theory of Games. Cambridge University Press

[16] Sirot E 2000. An evolutionarily stable strategy for aggressiveness in feeding groups. Behavioral Ecology 11, 351-356.

[17] Taylor PD and Jonker LB 1978. Evolutionary stable strategies and game dynamics. Math Bioscien 40, 145-156.

6 Appendix

A.1. Preliminaries

We first state some technical lemmas necessary later, although sometimes they are used only tacitly.

Lemma 6.1 (Garay et al. 2017, Lemma 2, p. 7) *The following system of nonlinear equations in n variables,*

$$u_i = \frac{1}{1 + \sum_{j=1}^{n} c_{ij} u_j}, \qquad 1 \le i \le n, \tag{24}$$

where the coefficients c_{ij} are positive numbers, has a unique solution in the unit hypercube $[0,1]^n$.

We also claim the following.

Lemma 6.2 *The solution $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in [0,1]^n$ of (24)*

  (i) *is a continuous function in*

$$\mathbf{c} := \left(c_{11}, \ldots, c_{1n}, c_{21}, \ldots, c_{2n}, \ldots, c_{n1}, \ldots, c_{nn} \right) \in \mathbb{R}_{\ge 0}^{n^2},$$

  (ii) *has positive coordinates uniformly separated from zero, namely*

$$\frac{1}{1 + \Sigma_{ij} c_{ij}} \le u_l \le 1 \qquad (1 \le l \le n).$$

Proof. To the proof of (i), assume that the sequence

$$\mathbf{c}_k = \left(c_{11}^{(k)}, \ldots, c_{1n}^{(k)}, c_{21}^{(k)}, \ldots, c_{2n}^{(k)}, \ldots, c_{n1}^{(k)}, \ldots, c_{nn}^{(k)} \right) \to \mathbf{c}.$$

For each positive integer k, let $\mathbf{u}_k = (u_1^{(k)}, u_2^{(k)}, \ldots, u_n^{(k)})$ be the unique solution of the system of equations

$$u_i = \frac{1}{1 + \sum_{j=1}^{n} c_{ij}^{(k)} u_j}, \qquad 1 \le i \le n,$$

in the unit hypercube $[0,1]^n$. As $\mathbf{u}_k \in [0,1]^n$ it is a bounded sequence. To see that $\mathbf{u}_k$ tends to $\mathbf{u}$ it, therefore, suffices to prove that any $\mathbf{u}_{k_l}$ convergent subsequence of $\mathbf{u}_k$ tends to the same $\mathbf{u}$. Denote by $\hat{\mathbf{u}}$ the limit of $\mathbf{u}_{k_l}$. Since

$$\frac{1}{1 + \sum_{j=1}^{n} b_{ij} x_j} - x_i$$

is a continuous function in $(\mathbf{b}, \mathbf{x})$ (consider it as a function of $n^2 + n$ variables on $\mathbb{R}_{\geq 0}^{n^2} \times [0,1]^n$), it follows that

$$0 = \frac{1}{1 + \sum_{j=1}^{n} c_{ij}^{(k_l)} u_j^{(k_l)}} - u_i^{(k_l)} \to \frac{1}{1 + \sum_{j=1}^{n} c_{ij} \hat{u}_j} - \hat{u}_i$$

which implies that

$$0 = \frac{1}{1 + \sum_{j=1}^{n} c_{ij} \hat{u}_j} - \hat{u}_i \quad \text{or, equivalently,} \quad \hat{u}_i = \frac{1}{1 + \sum_{j=1}^{n} c_{ij} \hat{u}_j}.$$

Because of the uniqueness of the solution, the limit $\hat{\mathbf{u}}$ must be the same for any convergent subsequent of $\mathbf{u}^{(k)}$ and $\hat{\mathbf{u}}$ must be equal to $\mathbf{u}$. Since $\mathbf{u}^{(k)}$ tends to this unique solution in $[0,1]^n$, the solution is continuous in $\mathbf{c}$.

The lower estimate in (ii) is apparent from (24) because $u_i \leq 1$ for every $1 \leq i \leq n$. $\blacksquare$

Corollary 6.3 *Consider a population of types $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_m \in S_n$ with frequencies $x_0, x_1, \ldots, x_m$.*

(i) *The active part ϱ_i of the different types given by the solution of the system:*

$$\varrho_i = \frac{1}{1 + \mathbf{p}_i T[\sum_{j=0}^{m} x_j \varrho_j \mathbf{p}_j]}$$

continuously depends on $\mathbf{x} = (x_0, x_1, \ldots, x_m)$. We use the notation $\varrho_i(\mathbf{x}) = \varrho_i(\mathbf{x}, \mathbf{p}_0, \ldots, \mathbf{p}_m)$ in this sense.

(ii) *Let*

$$\bar{\varrho}(\mathbf{x}) = \bar{\varrho}(\mathbf{x}, \mathbf{p}_0, \ldots, \mathbf{p}_m) := \sum_{i=0}^{m} x_i \varrho_i(\mathbf{x}, \mathbf{p}_0, \ldots, \mathbf{p}_m)$$

and

$$\bar{\mathbf{h}}(\mathbf{x}) = \bar{\mathbf{h}}(\mathbf{x}, \mathbf{p}_0, \ldots, \mathbf{p}_m) := \sum_{i=0}^{m} \frac{x_i \varrho_i(\mathbf{x}, \mathbf{p}_0, \ldots, \mathbf{p}_m)}{\bar{\varrho}(\mathbf{x}, \mathbf{p}_0, \ldots, \mathbf{p}_m)} \mathbf{p}_i.$$

24

If $\mathbf{y}$ is another frequency distribution such that $\bar{\mathbf{h}}(\mathbf{y}) = \bar{\mathbf{h}}(\mathbf{x}) =: \bar{\mathbf{h}}$, then both $\bar{\varrho}(\mathbf{x}) = \bar{\varrho}(\mathbf{y})$ and $\varrho_i(\mathbf{x}) = \varrho_i(\mathbf{y})$ $(0 \le i \le m)$.

(iii) If $\bar{\mathbf{h}}$ can be uniquely represented as a convex combination of $\mathbf{p}_0, \ldots, \mathbf{p}_m$[14] then x_i must be equal to y_i for every i.

(iv) If $\mathbf{p}$ is a convex combination of $\mathbf{p}_0, \ldots, \mathbf{p}_m$ then there is an $\mathbf{x} = (x_0, x_1, \ldots, x_m) \in S_{m+1}$ such that $\bar{\mathbf{h}}(\mathbf{x}) = \mathbf{p}$. Namely, if $\mathbf{p} = \sum_{i=0}^{m} \alpha_i \mathbf{p}_i$ then

$$x_i = \frac{\rho_{\mathbf{p}}}{\rho_i(\mathbf{p})} \alpha_i$$

where $\rho_{\mathbf{p}}$ is the unique solution in $[0,1]$ to the equation

$$\rho = \frac{1}{1 + \mathbf{p}T\rho\mathbf{p}} \tag{25}$$

and $\rho_i(\mathbf{p})$ denotes the expression

$$\frac{1}{1 + \mathbf{p}_i T \rho_{\mathbf{p}} \mathbf{p}} \qquad 0 \le i \le m. \tag{26}$$

Proof.

(i) The continuity of ϱ_i in $\mathbf{x}$ immediately follows from Lemma 6.2. (Set c_{ij} equal to $\mathbf{p}_i T x_j \mathbf{p}_j$.)

(ii) Since $\bar{\varrho}(\mathbf{x}) = (1-x)\varrho_0(\mathbf{x}) + x\tilde{\varrho}(\mathbf{x})$, Proposition 3.1 shows that

$$\bar{\varrho}(\mathbf{x}) = \frac{1}{1 + \bar{\mathbf{h}}T\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}}$$

and

$$\bar{\varrho}(\mathbf{y}) = \frac{1}{1 + \bar{\mathbf{h}}T\bar{\varrho}(\mathbf{y})\bar{\mathbf{h}}}$$

hold (as if the population consisted of only $\bar{\mathbf{h}}$-strategists). Lemma 6.1 says that the equation

$$\bar{\varrho} = \frac{1}{1 + \bar{\mathbf{h}}T\bar{\varrho}\bar{\mathbf{h}}}$$

---

[14]This is always true in the two cases (i) $m = 1$ and $\mathbf{p}_0 \ne \mathbf{p}_1$ and (ii) $\mathbf{p}_0, \ldots, \mathbf{p}_m$ are distinct pure strategies.

(with $\bar{\varrho}$ as the unknown) has a unique solution in $[0, 1]$ which implies at once that $\bar{\varrho}(\mathbf{x}) = \bar{\varrho}(\mathbf{y})$. Also,

$$\varrho_i = \frac{1}{1 + \mathbf{p}_i T \sum_j x_j \varrho_j \mathbf{p}_j} = \frac{1}{1 + \mathbf{p}_i T \bar{\varrho} \bar{\mathbf{h}}}$$

has a unique solution (in the unknowns ϱ_i) in $[0, 1]^{n+1}$ for every i from which $\varrho_i(\mathbf{x}) = \varrho_i(\mathbf{y})$ follows for every $0 \leq i \leq m$.

(iii) This is an immediate consequence of (ii) by comparing the coefficients in the representations $\bar{\mathbf{h}}(\mathbf{x})$ and $\bar{\mathbf{h}}(\mathbf{y})$.

(iv) Take the frequencies

$$x_i = \frac{\rho_{\mathbf{p}}}{\rho_i(\mathbf{p})} \alpha_i.$$

Then, by (25) and (26), we have that

$$\sum_{i=0}^{m} x_i = \sum_{i=0}^{m} \frac{\rho_{\mathbf{p}}}{\rho_i(\mathbf{p})} \alpha_i = \sum_{i=0}^{m} \alpha_i \frac{1 + \mathbf{p}_i T \rho_{\mathbf{p}} \mathbf{p}}{1 + \mathbf{p} T \rho_{\mathbf{p}} \mathbf{p}} = 1,$$

that is $\mathbf{x} = (x_0, x_1, \ldots, x_m)$ is a frequency distribution. Consider the polymorphic population of strategies $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_m$ with this frequency distribution $\mathbf{x}$. Then, $\rho_0(\mathbf{p}), \rho_1(\mathbf{p}), \ldots, \rho_m(\mathbf{p})$ satisfy the following system of equations corresponding the system (5) with $\mathbf{p}^* = \mathbf{p}_0$:

$$\varrho_i = \frac{1}{1 + \mathbf{p}_i T \left[\sum_{j=0}^{m} x_j \varrho_j \mathbf{p}_j \right]}. \tag{27}$$

Indeed, by a simple replacement, we get that

$$\frac{1}{1 + \mathbf{p}_i T \left[\sum_{j=0}^{m} x_j \rho_j(\mathbf{p}) \mathbf{p}_j \right]} = \frac{1}{1 + \mathbf{p}_i T \left[\sum_{j=0}^{m} \alpha_j \frac{\rho_{\mathbf{p}}}{\rho_j(\mathbf{p})} \rho_j(\mathbf{p}) \mathbf{p}_j \right]}$$

$$= \frac{1}{1 + \mathbf{p}_i T \left[\sum_{j=0}^{m} \rho_{\mathbf{p}} \alpha_j \mathbf{p}_j \right]} = \frac{1}{1 + \mathbf{p}_i T \rho_{\mathbf{p}} \mathbf{p}} = \rho_i(\mathbf{p})$$

On the other hand, $\varrho_0(\mathbf{x}), \varrho_1(\mathbf{x}), \ldots, \varrho_n(\mathbf{x})$ are also the solutions of the system of equations (27). By the uniqueness (see Lemma 6.1 in A.1.) we have that $\varrho_i(\mathbf{x}) = \rho_i(\mathbf{p})$. Consequently, we have that

$$\bar{\varrho}(\mathbf{x}) = \sum_{i=0}^{m} x_i \varrho_i(\mathbf{x}) = \sum_{i=0}^{m} x_i \rho_i(\mathbf{p}) = \sum_{i=0}^{m} \left(\frac{\rho_{\mathbf{p}}}{\rho_i(\mathbf{p})} \alpha_i \right) \rho_i(\mathbf{p}) = \rho_{\mathbf{p}} \sum_{i=0}^{m} \alpha_i = \rho_{\mathbf{p}}$$

26

and

$$\bar{\mathbf{h}}(\mathbf{x}) = \frac{1}{\bar{\varrho}(\mathbf{x})} \sum_{i=0}^{m} x_i \varrho_i(\mathbf{x}) \mathbf{p}_i = \frac{1}{\rho_{\mathbf{p}}} \sum_{i=0}^{m} \left(\frac{\rho_{\mathbf{p}}}{\rho_i(\mathbf{p})} \alpha_i \right) \rho_i(\mathbf{p}) \mathbf{p}_i = \sum_{i=0}^{m} \alpha_i \mathbf{p}_i = \mathbf{p}.$$

■

Lemma 6.4 *Let* $\mathbf{p}, \mathbf{r} \in S_2$. *Denote by* $\varrho_{\mathbf{p}}(\varepsilon), \varrho_{\mathbf{r}}(\varepsilon)$ *the unique solution in* $[0,1]^2$ *of the system:*

$$\varrho_{\mathbf{p}} = \frac{1}{1 + \mathbf{p}T[(1-\varepsilon)\varrho_{\mathbf{p}}\mathbf{p} + \varepsilon\varrho_{\mathbf{r}}\mathbf{r}]},$$

$$\varrho_{\mathbf{r}} = \frac{1}{1 + \mathbf{r}T[(1-\varepsilon)\varrho_{\mathbf{p}}\mathbf{p} + \varepsilon\varrho_{\mathbf{r}}\mathbf{r}]}.$$

Furthermore, $\bar{\varrho}(\varepsilon) := (1-\varepsilon)\varrho_{\mathbf{p}}(\varepsilon) + \varepsilon\varrho_{\mathbf{r}}(\varepsilon)$ *and*

$$\mathbf{q}(\varepsilon) := \frac{1}{\bar{\varrho}(\varepsilon)}[(1-\varepsilon)\varrho_{\mathbf{p}}(\varepsilon)\mathbf{p} + \varepsilon\varrho_{\mathbf{r}}(\varepsilon)\mathbf{r}].$$

Then $\mathbf{q}(0) = \mathbf{p}$, $\mathbf{q}(1) = \mathbf{r}$ *and* $\mathbf{q}(\varepsilon)$ *uniquely runs through the line segment between* $\mathbf{p}$ *and* $\mathbf{r}$ *as* ε *runs from 0 to 1 in such a way that* $0 \le \varepsilon_1 < \varepsilon_2 \le 1$ *implies that* $||\mathbf{q}(\varepsilon_1) - \mathbf{p}|| < ||\mathbf{q}(\varepsilon_2) - \mathbf{p}||$.

Proof. Corollary 6.3 (i) with the choice $\mathbf{p}_0 = \mathbf{p}$ and $\mathbf{p}_1 = \mathbf{r}$ shows that $\varrho_{\mathbf{p}}(\varepsilon)$ and $\varrho_{\mathbf{r}}(\varepsilon)$ are continuous in $\varepsilon \in [0,1]$. Therefore, both $\bar{\varrho}(\varepsilon)$ and $\mathbf{q}(\varepsilon)$ are continuous in ε. Since $\mathbf{q}(0) = \mathbf{p}$ and $\mathbf{q}(1) = \mathbf{q}$ the Bolzano-Darboux property of continuous function (intermediate value theorem) ensures that $\mathbf{q}(\varepsilon)$ runs through the line segment between $\mathbf{p}$ and $\mathbf{r}$ as ε runs from 0 to 1. Furthermore, by Corollary 6.3 (iii), $\mathbf{q}(\varepsilon_1) \ne \mathbf{q}(\varepsilon_2)$ also holds. Since $\mathbf{q}(0) = \mathbf{p}$, it follows that $||\mathbf{q}(\varepsilon_1) - \mathbf{p}|| < ||\mathbf{q}(\varepsilon_2) - \mathbf{p}||$. ■

A.2.

Lemma 6.5 *Let* $\mathbf{p}^* \in S_2$ *and* $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2) \in S_2$ *be the unique solution of* $\bar{\mathbf{h}}(\mathbf{x}) = \mathbf{p}^*$. *Then* $\mathbf{p}^*$ *is an ESS if and only if there is a* $\delta > 0$ *such that*

$$\hat{x}_1 W_1(\mathbf{x}) + \hat{x}_2 W_2(\mathbf{x}) > x_1 W_1(\mathbf{x}) + x_2 W_2(\mathbf{x}) = \bar{W}(\mathbf{x})$$

whenever $|x_1 - \hat{x}_1| < \delta$.

Proof. Consider a $\mathbf{p}^* \in S_2$. Without loss of generality, it can be assumed that $\mathbf{p}^* \neq \mathbf{e}_2$. We use the following notations:

$$\bar{\varrho}(\mathbf{x}) := x_1 \varrho_1(\mathbf{x}) + x_2 \varrho_2(\mathbf{x}) \quad \text{and} \quad \bar{\mathbf{h}}(\mathbf{x}) := x_1 \frac{\varrho_1(\mathbf{x})}{\bar{\varrho}(\mathbf{x})} \mathbf{e}_1 + x_2 \frac{\varrho_2(\mathbf{x})}{\bar{\varrho}(\mathbf{x})} \mathbf{e}_2. \quad (28)$$

Similarly, we introduce the notations $\hat{\varrho}(\mathbf{x})$ and $\hat{\mathbf{h}}(\mathbf{x})$, respectively, as follows:

$$\hat{\varrho}(\mathbf{x}) := \hat{x}_1 \varrho_1(\mathbf{x}) + \hat{x}_2 \varrho_2(\mathbf{x}) \quad \text{and} \quad \hat{\mathbf{h}}(\mathbf{x}) := \hat{x}_1 \frac{\varrho_1(\mathbf{x})}{\hat{\varrho}(\mathbf{x})} \mathbf{e}_1 + \hat{x}_2 \frac{\varrho_2(\mathbf{x})}{\hat{\varrho}(\mathbf{x})} \mathbf{e}_2. \quad (29)$$

Assume that $\hat{x}_2 < x_2$ (the case $x_2 < \hat{x}_2$ can be handled in the same way). It is easy to check this is equivalent to the inequality

$$\hat{x}_2 \frac{\varrho_2(\mathbf{x})}{\hat{\varrho}(\mathbf{x})} < x_2 \frac{\varrho_2(\mathbf{x})}{\bar{\varrho}(\mathbf{x})}$$

which implies that $\bar{\mathbf{h}}(\mathbf{x})$ lies on the line segment between $\hat{\mathbf{h}}(\mathbf{x})$ and $\mathbf{e}_2$. By Lemma 6.4, it is also true that $\bar{\mathbf{h}}(\mathbf{x})$ is located on the line segment between $\mathbf{p}^*$ and $\mathbf{e}_2$. Thus, there is an $\hat{\eta} = \hat{\eta}(\mathbf{x})$ and an $\eta = \eta(\mathbf{x})$, respectively, between 0 and 1, such that

$$\bar{\mathbf{h}}(\mathbf{x}) = (1 - \hat{\eta}) \frac{\rho_{\hat{\mathbf{h}}(\mathbf{x})}(\hat{\mathbf{h}}(\mathbf{x}), \mathbf{e}_2, \hat{\eta})}{\bar{\rho}(\hat{\mathbf{h}}(\mathbf{x}), \mathbf{e}_2, \hat{\eta})} \hat{\mathbf{h}}(\mathbf{x}) + \hat{\eta} \frac{\rho_{\mathbf{e}_2}(\hat{\mathbf{h}}(\mathbf{x}), \mathbf{e}_2, \hat{\eta})}{\bar{\rho}(\hat{\mathbf{h}}(\mathbf{x}), \mathbf{e}_2, \hat{\eta})} \mathbf{e}_2 \quad (30)$$

and

$$\bar{\mathbf{h}}(\mathbf{x}) = (1 - \eta) \frac{\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{e}_2, \eta)}{\bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta)} \mathbf{p}^* + \eta \frac{\rho_{\mathbf{e}_2}(\mathbf{p}^*, \mathbf{e}_2, \eta)}{\bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta)} \mathbf{e}_2, \quad (31)$$

respectively. Observe that $\bar{\rho}(\hat{\mathbf{h}}(\mathbf{x}), \mathbf{e}_2, \hat{\eta}) = \bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta) = \bar{\varrho}(\mathbf{x})$ and $\rho_{\mathbf{e}_2}(\hat{\mathbf{h}}(\mathbf{x}), \mathbf{e}_2, \hat{\eta}) = \rho_{\mathbf{e}_2}(\mathbf{p}^*, \mathbf{e}_2, \eta) = \varrho_2(\mathbf{x}).$[15] Note that η and x_2 mutually determine each other so we can write $\eta = \eta(x_2)$ or $x_2 = x_2(\eta)$ depending on what is given. A similar observation is true for the relationship between $\hat{\eta}$ and x_2.

Suppose, for some x_2, we have

$$\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{e}_2, \eta) \mathbf{p}^* A \underbrace{\bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta) \left[(1 - \eta) \frac{\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{e}_2, \eta)}{\bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta)} \mathbf{p}^* + \eta \frac{\rho_{\mathbf{e}_2}(\mathbf{p}^*, \mathbf{e}_2, \eta)}{\bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta)} \mathbf{e}_2 \right]}_{\bar{\varrho}(\mathbf{x}) \bar{\mathbf{h}}(\mathbf{x})}$$

$$> \rho_{\mathbf{e}_2}(\mathbf{p}^*, \mathbf{e}_2, \eta) \mathbf{e}_2 A \bar{\varrho}(\mathbf{x}) \bar{\mathbf{h}}(\mathbf{x}) = \varrho_2(\mathbf{x}) \mathbf{e}_2 A \bar{\varrho}(\mathbf{x}) \bar{\mathbf{h}}(\mathbf{x}). \quad (32)$$

---

[15]For example, to see that $\bar{\rho}(\hat{\mathbf{h}}(\mathbf{x}), \mathbf{e}_2, \hat{\eta}) = \bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta) = \bar{\varrho}(\mathbf{x})$ it is enough to observe that they satisfy the equation $\varrho = 1/[1 + \bar{\mathbf{h}}(\mathbf{x})T\varrho\bar{\mathbf{h}}(\mathbf{x})]$ with ϱ as the unknown and use Lemma 6.1 about uniqueness.

Multiply both sides of inequality (32) by $(1-\eta)$ then add $\eta\varrho_2(\mathbf{x})\mathbf{e}_2A\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x})$ to get that

$$\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x})A\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x}) > \varrho_2(\mathbf{x})\mathbf{e}_2A\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x}) \tag{33}$$

By subtracting $\hat{\eta}\varrho_2(\mathbf{x})\mathbf{e}_2A\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x})$ from both sides of (33), simplifying by $(1-\hat{\eta})$ and using (30) we obtain inequality:

$$\rho_{\hat{\mathbf{h}}(\mathbf{x})}(\hat{\mathbf{h}}(\mathbf{x}),\mathbf{e}_2,\hat{\eta})\hat{\mathbf{h}}(\mathbf{x})A\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x}) > \varrho_2(\mathbf{x})\mathbf{e}_2A\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x}). \tag{34}$$

Multiplying (34) by $\hat{\eta}$ and adding $(1-\hat{\eta})\rho_{\hat{\mathbf{h}}(\mathbf{x})}(\hat{\mathbf{h}}(\mathbf{x}),\mathbf{e}_2,\hat{\eta})\hat{\mathbf{h}}(\mathbf{x})A\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x})$ to both sides results in

$$\rho_{\hat{\mathbf{h}}(\mathbf{x})}(\hat{\mathbf{h}}(\mathbf{x}),\mathbf{e}_2,\hat{\eta})\hat{\mathbf{h}}(\mathbf{x})A\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x}) > \bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x})A\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x}) \tag{35}$$

where the right-hand side is just equal to $x_1W_1(\mathbf{x}) + x_2W_2(\mathbf{x})$. As regards the left-hand side, observe that, by Lemma 6.1, $\rho_{\hat{\mathbf{h}}(\mathbf{x})}(\hat{\mathbf{h}}(\mathbf{x}),\mathbf{e}_2,\hat{\eta}) = \hat{\varrho}(\mathbf{x})$.[16] Therefore the left-hand side just equals $\hat{x}_1W_1(\mathbf{x}) + \hat{x}_2W_2(\mathbf{x})$. The above reasoning shows the equivalence of inequalities (32), (33), (34) and (35) which means that

$$\hat{x}_1W_1(\mathbf{x}) + \hat{x}_2W_2(\mathbf{x}) > x_1W_1(\mathbf{x}) + x_2W_2(\mathbf{x}) \tag{36}$$

holds if and only if inequality (32) does.

If $\mathbf{p}^*$ is an ESS then consider ε_0 from Definition 2.1. By continuity (see Corollary 6.3), there exists a $\delta > 0$ such that if $0 < |x_2 - \hat{x}_2| < \delta$ then $0 < \eta(x_2) < \varepsilon_0$. Therefore, (32) holds which, as we have just seen, is equivalent to (36).

Conversely, assume the existence of a $\delta > 0$ such that (36) holds whenever $0 < |x_2 - \hat{x}_2| < \delta$. By Corollary 6.3 and Lemma 6.4, there is an $\varepsilon_0(\mathbf{e}_2)$ such that $0 < |x_2(\eta) - \hat{x}_2| < \delta$ whenever $0 < \eta < \varepsilon_0$.[17] Consider the strategy $\mathbf{r} \in S_2$ with $r_2 > p_2^*$ (the case $r_2 < p_2^*$ can be treated in a similar way). Then $\mathbf{r}$ is on the segment between $\mathbf{p}^*$ and

$$\mathbf{h}(\mathbf{e}_2,\varepsilon_0(\mathbf{e}_2)) := (1-\varepsilon_0(\mathbf{e}_2))\frac{\rho_{\mathbf{p}^*}(\mathbf{p}^*,\mathbf{e}_2,\varepsilon_0(\mathbf{e}_2))}{\bar{\rho}(\mathbf{p}^*,\mathbf{e}_2,\varepsilon_0(\mathbf{e}_2))}\mathbf{p}^* + \varepsilon_0(\mathbf{e}_2)\frac{\rho_{\mathbf{e}_2}(\mathbf{p}^*,\mathbf{e}_2,\varepsilon_0(\mathbf{e}_2))}{\bar{\rho}(\mathbf{p}^*,\mathbf{e}_2,\varepsilon_0(\mathbf{e}_2))}\mathbf{e}_2$$

---

[16]Both of them are equal to $1/[1 + \hat{\mathbf{h}}(\mathbf{x})T\bar{\varrho}(\mathbf{x})\bar{\mathbf{h}}(\mathbf{x})]$ as can be seen by mimicking the calculation from (11) to (13).

[17]If $\mathbf{p}^* \neq \mathbf{e}_1$ we can similarly find an appropriate $\varepsilon_0(\mathbf{e}_1)$ and if $\mathbf{p}^* = \mathbf{e}_1$ then all other strategies lie on the segment between $\mathbf{p}^*$ and $\mathbf{e}_2$, so we should only deal with the "side" of $\mathbf{p}^*$ between $\mathbf{p}^*$ and $\mathbf{e}_2$.

or on the segment between $\mathbf{h}(\mathbf{e}_2, \varepsilon_0(\mathbf{e}_2))$ and $\mathbf{e}_2$. In the former case set $\varepsilon_0(\mathbf{r})$ to be 1, in the latter one set $\varepsilon_0(\mathbf{r})$ to be the unique ζ_0 for which $\mathbf{h}(\mathbf{e}_2, \varepsilon_0(\mathbf{e}_2)) = \mathbf{h}(\mathbf{r}, \zeta_0)$ (such ζ_0 exists by Lemma 6.4). Then, again by Lemma 6.4, for any $0 < \zeta < \varepsilon_0(\mathbf{r})$ there is an $0 < \eta < \varepsilon_0(\mathbf{e}_2)$ with $\mathbf{h}(\mathbf{e}_2, \eta) = \mathbf{h}(\mathbf{r}, \zeta)$. Following the reasoning in the proof of Proposition 3.1 we see that $\bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta) = \bar{\rho}(\mathbf{p}^*, \mathbf{r}, \zeta)$ and, hence, $\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{e}_2, \eta) = \rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{r}, \zeta)$. From these observations and the argument above that inequalities (32), (33), (34) and (35) are equivalent, we conclude that the ensuing inequalities are equivalent with each other:

$$\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{r}, \zeta)\mathbf{p}^*A \underbrace{\left[(1-\zeta)\frac{\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{r}, \zeta)}{\bar{\rho}(\mathbf{p}^*, \mathbf{r}, \zeta)}\mathbf{p}^* + \zeta\frac{\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{r}, \zeta)}{\bar{\rho}(\mathbf{p}^*, \mathbf{r}, \zeta)}\mathbf{r} \right]}_{\bar{\rho}(\mathbf{p}^*,\mathbf{r},\zeta)\mathbf{h}(\mathbf{r},\zeta)} >$$

$$(37)$$

$$> \rho_{\mathbf{r}}(\mathbf{p}^*, \mathbf{r}, \zeta)\mathbf{r}A\bar{\rho}(\mathbf{p}^*, \mathbf{r}, \zeta)\mathbf{h}(\mathbf{r}, \zeta)$$
$$\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{r}, \zeta)\mathbf{p}^*A\bar{\rho}(\mathbf{p}^*, \mathbf{r}, \zeta)\mathbf{h}(\mathbf{r}, \zeta) > \bar{\rho}(\mathbf{p}^*, \mathbf{r}, \zeta)\mathbf{h}(\mathbf{r}, \zeta)A\bar{\rho}(\mathbf{p}^*, \mathbf{r}, \zeta)\mathbf{h}(\mathbf{r}, \zeta)$$
$$\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{e}_2, \eta)\mathbf{p}^*A\bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta)\mathbf{h}(\mathbf{e}_2, \eta) > \bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta)\mathbf{h}(\mathbf{e}_2, \eta)A\bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta)\mathbf{h}(\mathbf{e}_2, \eta)$$
$$\rho_{\mathbf{p}^*}(\mathbf{p}^*, \mathbf{e}_2, \eta)\mathbf{p}^*A\bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta)\mathbf{h}(\mathbf{e}_2, \eta) > \rho_{\mathbf{e}_2}(\mathbf{p}^*, \mathbf{e}_2, \eta)\mathbf{e}_2A\bar{\rho}(\mathbf{p}^*, \mathbf{e}_2, \eta)\mathbf{h}(\mathbf{e}_2, \eta).$$

We have seen that the last inequality is true for any $0 < \eta < \varepsilon_0(\mathbf{e}_2)$ whenever (36) holds for $(\hat{x}_1, \hat{x}_2)$. This proves that $\mathbf{p}^*$ is an ESS. ∎