Kaarboe, Oddvar M.; Tieman, Alexander F.

Working Paper

Equilibrium Selection under Different Learning Modes in Supermodular Games

Tinbergen Institute Discussion Paper, No. 99-061/1

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Kaarboe, Oddvar M.; Tieman, Alexander F. (1999) : Equilibrium Selection under Different Learning Modes in Supermodular Games, Tinbergen Institute Discussion Paper, No. 99-061/1, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/85664

Equilibrium Selection under Different Learning Modes in Supermodular Games *

Oddvar M. Kaarbøe † and Alexander F. Tieman ‡

June 15, 1999

Abstract

We apply the dynamic stochastic framework proposed by recent evolutionary literature to the class of strict supermodular games when two simple behavior rules coexist in the population, imitation and myopic optimization. We assume that myopic optimizers are able to see how well their payoff does relative to what they can get in the stage game and therefore experiment more in low payoff states. A clear-cut equilibrium selection result is obtained: the payoff dominant equilibrium emerges as the unique long run equilibrium. Furthermore, the expected waiting time until the payoff dominant equilibrium is reached is relatively short, even in the limit as the population size grows large.

Keywords: *Evolution, Imitation, Myopic Optimization, Payoff Dominant Equilibrium.*
JEL-code: C7, C72, C73, D83, D84.

1. Introduction

This paper studies how the long run prediction of the evolutionary model of Kandori and Rob (1995) (hereafter KR) changes for the class of strict supermodular games when two simple behavioral rules coexist in the population, namely imitation and myopic optimization. KR predict the long run behavior of an adaptive adjustment process in which myopic and boundedly

*The authors would like to thank Gerard van der Laan and Sjur D. Flåm for comments. The paper also benefited from discussion following a presentation at the University of Copenhagen.

†Corresponding author. University of Bergen, Dept. Economics, Fosswinckelsg. 6, 5007 Bergen, Norway. Ph: +47-55589226, Fax: +47-55589210, E-mail: oddvar.kaarboe@econ.uib.no

‡Tinbergen Institute and Free University, Dept. Econometrics, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Ph: +31-20-4446022, Fax: +31-20-4446020, E-mail: XTieman@econ.vu.nl. URL http://www.econ.vu.nl/medewerkers/xtieman

rational players base their strategy in the stage game (their action) on the expected payoff from the current population state. This adjustment process is perturbed by mutations whose effect is to introduce a stochastic flow of players adopting any action. By using this framework KR show how evolutionary forces and mutations interact to pin down the set of equilibria that is most likely to be observed in the long run, assuming that players are repeatedly randomly matched in pairs to play a strict supermodular stage game. Specifically, they show how the geometry of the best response correspondence helps to identify the set of long run equilibria. One contribution of this paper is to demonstrate that when the two simple behavior rules, imitation and myopic optimization, coexist in the population, a much more clear-cut equilibrium selection result is obtained: the payoff dominant equilibrium (see Harsanyi and Selten (1988)) is selected as the unique long run equilibrium for the class of strict supermodular games. Furthermore we show that the expected waiting time until the long run equilibrium is reached is relatively short, even in the limit as the population size grows large. Hence convergence may in fact be rapid even though the mutation rate is small.

Our results do (of course) depend on the specific assumptions made on the behavior rules, imitation and myopic optimization. As in Vega-Redondo (1995), the different types of players (or different behavior rules) are interpreted as representing different degrees of sophistication. We call players with the lowest degree of sophistication imitators. They are fairly unsophisticated in the sense that they have no knowledge whatsoever about the stage game they are playing. When they imitate, they are implicitly hoping that the players they decide to mimic are already well adapted. For simplicity, we pose that imitators observe average payoff to each action played in the last period and copy the action that got the highest average payoff. Myopic optimizers (or best reply players) are somewhat more sophisticated. They know the payoffs in the stage game and their best-reply correspondence. They play a best response to the population state they observe, which can thus be seen as their expectation of next period's state, i.e. we assume that myopic optimizers believe their opponents to stay with their current actions. Hence, myopic optimizers form adaptive expectations. Furthermore, we pose that myopic optimizers, who are able to see how well their payoff does relative to what they can get in stage game, experiment more in situations in which they have realized a lower relative payoff, since they 'reason' that experimenting is potentially more rewarding in such a situation than in a situation in which high relative payoffs are realized. This way the mutation rate of myopic optimizers is state dependent. In contrast, imitators do not know the stage game and thus have a flat mutation rate over the different states. Alternatively, like in Gale and Rosenthal (1999), we may specify the two different types by how they learn: either a player

learns indirectly (i.e. copies another player and hopes that this player is well adapted), or she learns directly (i.e. looks around and tries to do the best she can). Furthermore, when a player learns directly, she evaluates her payoff relative to what she might get and experiments more in situations with lower relative payoffs. The former type of learning is imitation, the latter is what we call myopic optimization.

The motivation for this paper is threefold. First, the dynamic stochastic framework proposed by Kandori, Mailath, and Rob (1993) (KMR) and Young (1993) (together labelled as KMRY) focuses attention on rare mutations as an equilibrium selection device. They incorporate myopic and boundedly rational players who base the action they will play on the expected payoff from the current population state. Neglecting mutations, the population state evolves through a Darwinian[1] difference equation in KMR and through adaptive play in Young (1993). Basically, in KMRY players can be considered to play a best reply, given their limited information and bounded rationality. The important result KMRY obtain on 2×2 games is the selection of the risk dominant equilibrium as the unique long run equilibrium. In contrast, Robson and Vega-Redondo (1996) model players that imitate the strategy whose average payoff was highest in the previous period. This model yields selection of the payoff dominant equilibrium as the unique long run equilibrium in the class of symmetric 2×2 coordination games. Moreover, convergence is much faster in this model than in KMRY.[2]

However, non of these models provide explains why players use a specific update rule, like imitation or myopic optimization.Vega-Redondo (1995) does address this issue implicitly. He argues that when players update their expectations according to different update rules, any Nash equilibrium in the stage game can get positive measure in the invariant distribution when the mutation rate goes to zero. Vega-Redondo (1996), Section 6.7, studies a more specific version of this model, in which there explicitly is competition between imitators, myopic optimizers, and so called 'dynamic optimizers'. Dynamic optimizers are very sophisticated players. They form self-confirming expectations about next period's state and play a best-reply to these expectations. We define self-confirming expectations as expectations against which no evidence occurs as long as play remains on the equilibrium path. Along the equilibrium path self-confirming expectations are thus consistent with observed play. Thus playing a best-reply to self-confirming expectations yields maximal payoff as long as play remains on the equilibrium path. According to Vega-Redondo (1996), the most interesting results emerge when a narrow range of sophistication is studied, i.e. when only imitators and myopic optimizers compete.

[1] We define Darwinian dynamics in the same way as Kandori, Mailath, and Rob (1993).

[2] Robson and Vega-Redondo (1996) also extend these results to games of common interest.

A restriction of these models of Vega-Redondo is that they focus on symmetric coordination games only.

Our paper unifies the approaches taken above. It provides an evolutionary framework in which selection among the rules of behavior myopic optimization and imitation takes place for general $m \times m$ strict supermodular games. Furthermore, we believe that in general players are not sophisticated enough to form self-confirming expectations.[3] Therefore we focus on competition between the behavioral rules imitation and myopic optimization.

Second, the theory of supermodular games provides a framework for the analysis of systems marked by complementarities. This class of games, introduced by Topkis (1979) and further explored by Milgrom and Roberts (1990) and Vives (1990), includes models of oligopolistic competition, macroeconomic coordination failure, Bertrand price competition, bank runs, and R&D competition. Supermodular games are characterized by the fact that each player's action set is partially ordered, the marginal returns to increasing one's action rise with increases in the opponents' actions. In the case of multidimensional actions, the marginal returns to any one component of the player's action rise with increases in the other components. As a result, these games exhibits strategic complementarities that yield monotone increasing individual best responses, as the actions are completely ordered. Furthermore, as remarked by e.g. Milgrom and Roberts (1990), an analysis of supermodular games like ours is entirely ordinal in character, since it focuses on inequalities between payoffs. Thus, the payoffs in a supermodular stage game can be regarded as ordinal utility levels, which is often neglected in economic theory, as it tends to focus on Von Neumann-Morgenstern expected utility theory. We analyze strict supermodular games under the assumption that we restrict payoffs to be within certain bounds. That this is without loss of generality follows directly from the argument above on the ordinal nature of utility levels.

Finally, we follow the suggestion of Bergin and Lipman (1996) and incorporate an explicit model of the mutation process.[4] The striking result of Bergin and Lipman (1996) is that any invariant distribution of the 'mutationless' process can be obtained in a setting with mutations, as in the limit the probability of mutation approaches zero, if in the model the relative

[3] Note that the presence of dynamic optimizers can upset any selection result among symmetric Nash equilibria, through expectational drift (see Vega-Redondo (1995)). However, when we assume dynamic optimization to come with a (computational) cost, in any equilibrium, selection pressure will see to it that dynamic optimizers become extinct from a population.

[4] van Damme and Weibull (1998) also suggest a model of varying mutation rates. However, they focus attention on mistakes, while we focus attention on experimentation. Moreover, they restrict attention to 2×2 games. Robles (1998) also suggests a model of varying mutation rates.

probabilities of the actions to which a mutation can change a player's action can approach zero or infinity. Hence, to generate more precise equilibrium selection predictions, economically interesting conditions on mutation rates must be incorporated in the model. In particular, if mutations reflect experimentation (as in our model), and the game has a payoff dominant equilibrium (as in strict supermodular games), Bergin and Lipman (1996) (p. 944) suggest that one might expect experimentation to occur at a lower rate in the state where all players play according to the payoff dominant equilibrium than in any other states. Of course this only holds for players that are able to rank the payoffs in the stage game.

The structure of the paper is as follows. Section 2 provides the general model. In sections 3 and 4, we analyze convergence in the absence and presence of mutations. Finally, in Section 5, we discuss our results.

2. The Model

We consider a finite population $\mathcal{N} := \{1, 2, ..., N\}$ consisting of N players with N even. In each period $t = 1, 2, \ldots$ all players are randomly matched in pairs to play a stage game.

2.1. The Stage Game

We consider a one population model. Thus, there is no difference in role between being the row or the column player and therefore, the stage game is symmetric. The stage game we posit is a strict supermodular game. The set $\mathcal{M}$ of pure actions in a strict supermodular stage is by definition partially ordered. Here we follow KR and assume that $\mathcal{M}$ is finite and completely ordered from low (action 1) to high (action M), i.e. $\mathcal{M} := \{1, 2, ..., M\}$. A player's payoff when she and her match choose action m and m' respectively, is denoted by $u(m, m')$.

Definition 2.1. *The stage game is a strict supermodular game if for any pair of strategies* $1 \leq m < m' \leq M$ *the payoff differences* $u(m', m'') - u(m, m'')$ *are strictly increasing in* m''.

The ordering of the actions and the supermodularity of the stage game ensures that all diagonal payoffs in the stage game are rankable in the Pareto sense, i.e. for all $m, m' \in \mathcal{M}$, $1 \leq m < m' \leq M$, it holds that $u(m, m) < u(m', m')$. We constrain off-diagonal payoffs to action M, $u(M, m)$ and $u(m, M)$, with $m \neq M$, to exceed a certain lower bound, specified in Section 4.

The following proposition states some well known results for strict supermodular games.

Proposition 2.2. *(KR, Proposition 6) After all strictly dominated strategies have been iteratively removed from the game, then*
i) the smallest and largest strategies are pure action Nash Equilibria (NE for short),
ii) no asymmetric NE in pure actions exists, and
iii) for a generically selected supermodular game all pure action NE are strict.

Thus, in a symmetric strict supermodular game, the set of pure action NE is a subset of the main diagonal.

Definition 2.3. $\mathcal{M}^* := \{m^* \in \mathcal{M} \mid (m^*, m^*)$ *is a pure action NE}.*

2.2. The Players' Types and the State Space

At every time $t = 1, 2, \ldots$ each player in the population uses a particular update rule, which we refer to as a player's type. An update rule specifies how a player updates the action she plays in the stage game, when given the possibility to revise her action. We consider the update rules '*imitation*' (ι) and '*myopic optimization*' (μ) and thus we accommodate two types of players. Players that update according to ι are called imitators, while players that update according to μ are called myopic optimizers.

In every period t each player is characterized by a pair (m, i), $m \in \mathcal{M}$, $i \in \{\iota, \mu\}$, identifying the action m she currently plays and her type i. For every period t, the state $s(t) = (s_1^\iota(t), s_1^\mu(t), ..., s_M^\iota(t), s_M^\mu(t))$ is a vector of summary statistics, whose element, $s_m^i(t)$, $m \in \mathcal{M}$, $i = \iota, \mu$, represents the number of type i-players using action m at time t. Thus, the state space is given by $\mathcal{S} = \{1, 2, \ldots, N\}^{2M}$, where for every $s \in S$, we have that $\sum_{m \in \mathcal{M}} (s_m^\iota + s_m^\mu) = N$. We take account of the state $s(t)$ at the start of period t. The total number of players playing action m at time t is $s_m(t) = s_m^\iota(t) + s_m^\mu(t)$ and the number of myopic optimizers in the population at time t is $N_t^\mu = \sum_{m=1}^M s_m^\mu(t)$. The number of imitators at time t is $N_t^\iota = N - N_t^\mu$. We refer to the vector with entries $s_m(t)$ by $\widetilde{s}(t) = (s_m(t))_{m \in \mathcal{M}}$ and define the vector $s^{-n}(t) = (s_m^{-n}(t))_{m \in \mathcal{M}}$ as the vector representing the total number of players playing action m at time t, when player n, $n \in \mathcal{N}$, is excluded from the population and we denote the m-th entry of $s^{-n}(t)$ by $s_m^{-n}(t)$. Note that both $\widetilde{s}(t)$ and $s^{-n}(t)$ are frequency distributions and thus the concept of first order stochastic dominance between states is well defined.

As stated before, at each time t, all players are matched in pairs once to play the stage game. All possible matches[5] have the same probability of occurring.. By $g_t^i(m, m')$ we denote

[5] It can be shown by induction that there are $(N-1) \cdot (N-3) \cdot \ldots \cdot 3 \cdot 1$ possible matches.

the number of matches at time t between a player of type i, $i = \iota, \mu$, playing action m and a player playing action m'. Note that all matches m-m' are counted twice, namely once as a match of an m-player with an m'-player and once as a match of an m'-player with an m-player. Furthermore, we define $g_t(m, m') = g_t^{\iota}(m, m') + g_t^{\mu}(m, m')$. From the above it follows that $g_t(m, m)$ is twice the number of m-m matches. Since all players are matched once, we have that $s_m(t) = \sum_{m' \in \mathcal{M}} g_t(m, m')$. Now, let $\overline{u}_m(t)$ denote average payoff of action $m \in \mathcal{M}$, at time t. I.e.,

$$\overline{u}_m(t) := \frac{1}{s_m(t)} \sum_{m' \in \mathcal{M}} u(m, m') g_t(m, m') .$$

2.3. The Mutation-Free Dynamics

The mutation-free dynamics operate on two levels. At the first level, at each time t, some players get the possibility to revise their action. They do so according to a particular update rule. At the second level, at each time t, some players get the possibility to switch update rules, i.e. to update their types. We now specify both processes in detail.

2.3.1. Updating Actions

At every $t = 1, 2, ...$ and before play is conducted each player takes an independent draw from a Bernoulli trial. With probability $\upsilon \in (0, 1)$ this draw produces the outcome 'learn', and the player chooses the new action as dictated by her type. With the complementary probability $1 - \upsilon$, the draw produces the outcome 'do not learn' and the player stays with her action. The possibility to update is called a *learning draw*.

Assumption A (*on imitators*). Imitators only observe average payoff to each action played in the last period and imitate the action that performed best, (i.e. got the highest average payoff). In case of ties, there is a positive probability of choosing each of the best actions.

Assumption B (*on myopic optimizers*). Myopic optimizers observe the fraction of players playing each action in the last period and choose a myopic pure best reply to the state from which she herself is excluded. In case of ties, there is a positive probability of choosing each of the best actions.

Now we are able to define the update rules in terms of the state space. An imitator that gets the learning draw switches to action $m^* \in \arg\max_m \overline{u}_m(t)$. Thus an imitator that plays action m at time t and updates, contributes to a change in the population state by decreasing s_m^{ι} by 1 and increasing $s_{m^*}^{\iota}$ by 1. A myopic optimizer that is currently playing action m and

updates, switches to action $m^* \in \arg\max_m u(m, m')\frac{s_{m'}^{-n}(t)}{N-1}$. When m^* is not unique, both types pick an arbitrary action from their set of argmax-es. For both types it is possible that $m^* = m$, i.e. they remain with their current action when updating.

2.3.2. Updating Types

After the stage game has been played by all pairs of players, we calculate average payoffs among myopic optimizers and imitators. We label these values $\overline{u}_t^{\mu}$ and $\overline{u}_t^{\iota}$. Hence,

$$\overline{u}_t^i = \frac{1}{N_t^i} \sum_{m \in \mathcal{M}} \sum_{m' \in \mathcal{M}} u(m, m') g_t^i(m, m'), \qquad i = \iota, \mu.$$

We posit that with probability $\theta \in (0, 1)$ each player receives the opportunity of revising her type. In that case, she changes type if and only if the average payoff to the other type in the last period is strictly higher that the average payoff her own type received in the same period, i.e. a player of type i, $i = \iota, \mu$, switches type iff $\overline{u}_t^i < \overline{u}_t^j$, $j = \iota, \mu$, $j \neq i$. When a player changes type we assume that she starts to play an action picked at random from the set of actions played by players of her new type.

2.4. Mutation Dynamics

At both levels we allow the above described mutation-free dynamics to be slightly perturbed by deviations. We refer to these deviations as a birth & death process on the type level and as mutations on the action level. Before we specify both processes in details, we present figure 2.1, a graphic overview of the sequence of events during a time period t.

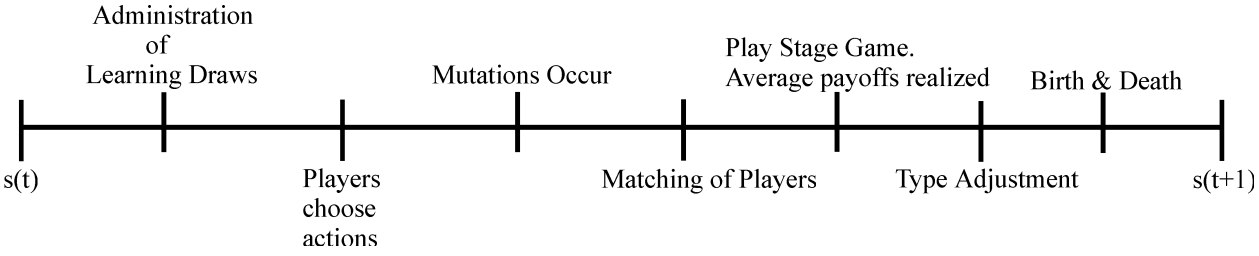

Figure 2.1: The sequence of events during a time period.

2.4.1. Birth & Death at the Type Level

The birth & death process sees to it that some randomly chosen players 'die' and are replaced by new players of either type. This event occurs after the type adjustment according to the

mutation-free dynamics has taken place.

We assume that at each time t there is a positive probability $1 > 2\delta \geq 0$ that a player dies. Setting $\delta = 0$ means that no birth & death takes place. When $\delta > 0$ and a player dies, she is replaced by a newborn player, who is either an imitator or a myopic optimizer, both with probability $\frac{1}{2}$. This formulation boils down to each player having a probability $\delta > 0$ of being replaced by a new player of a different type. Formally, let the random variables $V_t\left(\widetilde{N}_t^{\iota}, \delta\right)$ and $W_t\left(\widetilde{N}_t^{\mu}, \delta\right)$ denote the number of imitators and myopic optimizers respectively that switch type at time t and thus become myopic optimizers or imitators at time $t+1$. Note that $V_t\left(\widetilde{N}_t^{\iota}, \delta\right)$ and $W_t\left(\widetilde{N}_t^{\mu}, \delta\right)$ both have binomial distributions with parameters $n = \widetilde{N}_t^{\iota}$ and $n = \widetilde{N}_t^{\mu}$ respectively and $p = \delta$, where $\widetilde{N}_t^{i}$, is the number of type i players in the population at time t after players were able to revise their type according to the mutation-free type adjustment dynamics. Now, we have that

$$N_{t+1}^{\iota} = \widetilde{N}_t^{\iota} - V_t\left(\widetilde{N}_t^{\iota}, \delta\right) + W_t\left(\widetilde{N}_t^{\mu}, \delta\right)$$

and

$$N_{t+1}^{\mu} = \widetilde{N}_t^{\mu} + V_t\left(\widetilde{N}_t^{\iota}, \delta\right) - W_t\left(\widetilde{N}_t^{\mu}, \delta\right).$$

2.4.2. Mutations at the Action Level

At each time t, each player is subject to some common and independent (across players and time) probability of making a mistake in the implementation of her action. On top of that, myopic optimizers experiment with an independent (across players and time) probability. In both cases the player chooses any action in a purely arbitrarily manner. These mutations happened after action adjustments but prior to play being conducted. We interpret them as the joint result of mistakes and experimenting.

Assumption C (*on mistakes*). At every time t, each player makes a mistake in the implementation of her action with some common and independent probability $\varepsilon > 0$. In that case she plays an action $m \in \mathcal{M}$ with positive probability on each $m \in \mathcal{M}$.

Imitators are rather unsophisticated players that simply observe average payoff to each action played in the last period and imitate the one that performed best (Assumption A). They thus learn in an indirect way by copying another player's action, hoping that this player is well adapted. Hence, an imitator has no knowledge whatsoever concerning the structure of the game. Specifically, imitators do not even know the payoffs in the stage game. Thus, imitators only make mistakes and do not experiment.

Myopic optimizers are more sophisticated. They use a (myopic) best-reply to the last-period population state. So, presumably they know their best-reply correspondence. We assume that on top of their best-reply correspondence they also know the payoffs in the stage game and thus that they are able to evaluate their payoff relative to other payoffs in the stage game. We assume that myopic optimizers experiment less (more) when they receive relatively high (low) payoffs. In particular, since the stage game analyzed here has a payoff dominant equilibrium, experimentation occurs at a lower rate in the state where all players play according to the payoff dominant equilibrium than in any other equilibrium state. Note that also Bergin and Lipman (1996) (p. 944) argue for introducing state dependent mutations which reflect experimentation in a similar way.

Assumption D (*on experimentation*). The higher the payoff myopic optimizers receive, the lower their rate of experimentation.

For myopic optimizers, the combination of making mistakes in the implementation of their action and experimenting is modelled as a base rate reflecting the ε-probability of making a mistake and an additional state dependent mutation rate representing experimentation.

To operationalize assumptions D, we rank the payoffs that are obtainable in the stage game from low ($\widetilde{1}$) to high ($\widetilde{M^2}$). We introduce a strictly increasing function $0 < \beta(\cdot) < 1$ of these ranked payoffs ($\widetilde{1}$ to $\widetilde{M^2}$). The mutation rate of each myopic optimizer is taken to be $\varepsilon^{\beta(\cdot)} > \varepsilon$.

The composition of the adjustment processes generates a discrete-time Markov-process over the finite state space $\mathcal{S}$, whose transition matrix is denoted by $P(\varepsilon, \delta) = (p_{ss'}(\varepsilon, \delta))$. An element $p_{ss'}(\varepsilon, \delta)$ represents the transition probability of moving to state s' at time $t+1$ conditional on being in state s at time t. The dynamics without mutations at the action level and without birth & death corresponds to $P(0, 0)$.

The occurrence of mutations and birth & death implies that every transition has positive probability. It is a standard result that such Markov chains have a unique stationary (invariant) probability distribution. Let $\phi_\delta(\varepsilon)$ denote the unique invariant distribution of $P(\varepsilon, \delta)$ for each $\varepsilon > 0$ and fixed $\delta \geq 0$. Our aim is to characterize the unique invariant distribution $\phi_\delta^* := \lim_{\varepsilon \to 0} \phi_\delta(\varepsilon)$.[6] The states that have strict positive measure under ϕ_δ^* are called long run equilibria.

[6] Using arguments in Freidlin and Wentzell (1984), Young (1993) has shown that this limit exists.

3. Convergence to Limit Sets

As a fist step towards compution the set of long run equilibria we will identify the limit sets under the adjustment process without mutations. These are the supports of the invariant distributions of $P(0,0)$. First we consider a population consisting of only imitators. Second, we identify the limit sets in populations consisting of both imitators and myopic optimizers.

Proposition 3.1. *Consider the case in which all players are imitators. Then the only limit sets (of the mutation-free dynamics) are the singletons consisting of monomorphic states.*

Proof.
Obviously, every monomorphic state defines, as a singleton, a limit set. To see that no other limit sets exist, it is sufficient to show that starting from any other state, the process converges to a monomorphic state with positive probability.

Again let $m^* \in \arg\max_m \overline{u}_m(t)$. Take a polymorphic state $s(t)$, and let all players receive learning draw (on the action level). In period $t+1$, they will all switch to an action in $\arg\max_m \overline{u}_m(t)$. If m^* is unique, the state in period $t+1$ is by definition monomorphic. If not, by Assumption A, there is a positive probability that all players choose the same action, and the proposition is established. □

Before we proceed we state two propositions from KR. The first proposition is on monotonicity of best responses over the simplex of mixed actions, while the second proposition says that states which mimic mixed action equilibria are strongly unstable in the sense that they are not even stationary points of the best-response dynamic.

Proposition 3.2. *(KR, Proposition 7). Given two states s and s', with $(s')^{-n} \succ s^{-n}$, where $\succ$ refers to first-order stochastic-dominance. For myopic optimizer n, $n = 1, 2, \ldots, N$, let $br(s^{-n})$ denote the set of best responses to s^{-n}. Then $\min br\left((s')^{-n}\right) \geq \max br(s^{-n})$.*

Proposition 3.3. *(KR, Proposition 8). Suppose all myopic optimizers are taking best responses under s. Then in a strict supermodular game, only one action is played under s.*

We now state the convergence result for a population consisting of both imitators and myopic optimizers.

Proposition 3.4. *Consider the case in which the population consists of both imitators and myopic optimizers. Then, the limit sets of the mutation-free dynamics correspond one-to-one with the collection of pure strategy NE.*

The following lemma is useful in the proof of Proposition 3.4.

Lemma 3.5. *Suppose* $s = \{0, .., s^{\iota}_m, s^{\mu}_m, .., 0\}$, *with* $s^{\iota}_m + s^{\mu}_m = N$, *and that* $m' \in br(s^{-n})$ *for all myopic optimizers, where* $m' \neq m$. *If i)* $m' > m$ *then* $u(m', m) > u(m, m)$, *ii)* $m' < m$ *then* $u(m', m) > u(m, m) > u(m', m') > u(m, m')$.

Proof.
i) follows directly from the optimality of m' relative to s^{-n}, as is the case for the first inequality in ii). The second inequality in ii) follows from the Pareto rankability of the diagonal payoffs. Finally, the third inequality follows from the definition of a strict supermodularity stage game, and can be shown as follows.

From the strict supermodular structure of the stage game, we have that for $m' < m$,

$$u(m, m) - u(m', m) > u(m, m') - u(m', m').$$

Using $u(m', m) > u(m, m)$, we get

$$u(m, m) - u(m, m') > u(m', m) - u(m', m') > u(m, m) - u(m', m'),$$

and hence $u(m, m') < u(m', m')$. □

Proof (Proposition 3.4)**.**
Obviously, every pure (symmetric) Nash equilibrium, which is a monomorphic state that is a best reply to itself, is a limit set. To see that no other limit sets exist, it is sufficient to show that starting from any other state, the process converges to a pure strategy Nash equilibrium with positive probability.

The proof has two parts. First we will show that starting in any monomorphic state, a path that leads to a pure strategy Nash equilibrium with positive probability exists. Secondly, we show that starting in any polymorphic state, with positive probability, the system ends in a monomorphic state.

Part 1: Suppose the system is in a monomorphic state at time t. I.e., $s(t) = \{0, .., s^{\iota}_m, s^{\mu}_m, .., 0\}$, with $s^{\iota}_m + s^{\mu}_m = N$. (Remember that we identify the state at the beginning of each period.) Consider now the following possibilities: (of course, if $m \in br(s^{-n}(t))$, the state is a pure action Nash equilibrium, and we are done).

1A) $m' = \min\{\overline{m} | \overline{m} \in br(s^{-n}(t))\} > m$ and N^{μ} is even. Let all myopic optimizers learn and get matched with other myopic optimizers. Since $m' > m$ and because of the specific matching, $\overline{u}_{m'}(t) = u(m', m') > u(m, m) = \overline{u}_m(t)$. Suppose no type adjustments occurs. The state at

$t+1$ is $s(t+1) = \{0, ...s^{\iota}_{m} = N^{\iota}, \ldots, s^{\mu}_{m'} = N^{\mu}, ..., 0\}$. Now let all imitators learn. Since $\overline{u}_{m'}(t) > \overline{u}_{m}(t)$, they all change action from m to m'. Hence, the system is in a monomorphic state $s(t+2) = \{0, ..., s^{\iota}_{m'}, s^{\mu}_{m'}, ..., 0\}$, with $s^{\iota}_{m'} + s^{\mu}_{m'} = N$.

1B) $m' = \min\{\overline{m} | \overline{m} \in br(s^{-n}(t))\} > m$ and N^{μ} is odd. In this case let $(N^{\mu} - 1)$ myopic optimizers learn and get matched with one of the other myopic optimizers that just got a learning draw. Since $m' > m$ and because of the specific matching, $\overline{u}_{m'}(t) = u(m', m') > u(m, m) = \overline{u}_{m}(t)$. Furthermore, let one imitator receive the opportunity of adjusting her type. Since

$$\overline{u}^{\mu}_{t} = \frac{(N^{\mu}_{t} - 1)u(m', m') + u(m, m)}{N^{\mu}_{t}} > u(m, m) = \overline{u}^{\iota}_{t},$$

she becomes a myopic optimizer. Suppose she starts to play action m at $t+1$.[7] Now let all imitators learn. Since $\overline{u}_{m'}(t) > \overline{u}_{m}(t)$, they all change action from m to m'. The state at $t+1$ thus is $s(t+1) = \{0, ..., s^{\mu}_{m} = 2, \ldots, s^{\iota}_{m'}, s^{\mu}_{m'}, ..., 0\}$, with $s^{\iota}_{m'} + s^{\mu}_{m'} = N - 2$. Suppose the two myopic optimizers playing m get matched. Average type payoffs then are

$$\overline{u}^{\iota}_{t+1} = u(m', m') > \frac{(N^{\mu}_{t+1} - 2)u(m', m') + 2u(m, m)}{N^{\mu}_{t+1}} = \overline{u}^{\mu}_{t+1}.$$

Now, let the two myopic optimizers playing m receive the opportunity of adjusting their prior type. Since the imitators are earning the highest average payoff, the two players will change type and become imitators. Furthermore, they will start to play action m' in period $t+2$. Hence, the system is in a monomorphic state $s(t+2) = \{0, ..., s^{\iota}_{m'}, s^{\mu}_{m'}, ..., 0\}$, with $s^{\iota}_{m'} + s^{\mu}_{m'} = N$.

We have now shown that when $m' = \min\{\overline{m} | \overline{m} \in br(s^{-n}(t))\} > m$, the system moves to the monomorphic state with $s_{m'} = N$ with positive probability. If $m' \in \mathcal{M}^{*}$, by definition m' is a pure action Nash equilibrium. If not, hand all myopic optimizers the possibility to learn on the action level again. They now all face a state $s^{-n}(t+2) \succ s^{-n}(t)$, since $m' > m$. Hence, for each myopic optimizer n, $\min br\left(s^{-n}(t+2)\right) \geq \max br\left(s^{-n}(t)\right) \geq m'$, which was their minimal best reply in period t. (This fact follows from Proposition 3.2). Furthermore, since m' is not a Nash equilibrium, $\min br\left(s^{-n}(t+2)\right) \neq m'$, leading to $\min br\left(s^{-n}(t+2)\right) > m'$. By repeating the procedures given above, and using the fact that the state space is finite, we conclude that the system converges to a monomorphic state which is a pure strategy Nash equilibrium.

1C) $m' = \max\{\overline{m} | \overline{m} \in br(s^{-n}(t))\} < m$. Since m is a monomorphic state, all myopic optimizers choose the same best reply $m' < m$ with positive probability. From lemma 3.5 we now

[7] Remember that when a player changes her type, she starts to play one of the actions that her new type already is playing.

have the following payoff ranking

$$u(m',m) > u(m,m) > u(m',m') > u(m,m').$$

i) If $N^{\mu} \leq \frac{N}{2}$, let all myopic optimizers learn on the action level. With positive probability they all switch to playing action m'. Now, let them all get matched with imitators in period t. These matches give the following average payoff ranking: $\overline{u}_{m'}(t) = u(m',m) > u(m,m') = \overline{u}_{m}(t)$. Suppose no players receive the opportunity of revising their type in period t. The state in period $t+1$, is $s(t+1) = \{0,...,s^{\mu}_{m'} = N^{\mu}_{t},\ldots,s^{\iota}_{m} = N^{\iota}_{t},...,0\}$. Now, let no myopic optimizers and all imitators learn at the action level. Since $\overline{u}_{m'}(t) > \overline{u}_{m}(t)$, they start to play action m'. Further, let no player receive a learning draw on the type level. Hence, $s(t+2) = \{0,...,0,s^{\iota}_{m'},s^{\mu}_{m'},0,\ldots,0\}$, with $s^{\iota}_{m'} + s^{\mu}_{m'} = N$

ii) $N^{\mu}_{t} > \frac{N}{2}$. Let $\frac{N}{2}$ myopic optimizers learn. With positive probability, they all switch to action $m' < m$. Let these myopic optimizers be matched with the $\frac{N}{2}$ players who still play action m. From lemma 3.5, $\overline{u}_{m'}(t) = u(m',m) > u(m,m') = \overline{u}_{m}(t)$. Let no players change type in period t. The state at $t+1$ is $s(t+1) = \{0,...,s^{\mu}_{m'} = \frac{N}{2},\ldots,s^{\mu}_{m} = \frac{N}{2} - N^{\iota}_{t}, s^{\iota}_{m} = N^{\iota}_{t},...,0\}$. Let all imitators learn at the action level in period $t+1$. Since $\overline{u}_{m'}(t) > \overline{u}_{m}(t)$, they change action from m to m'. Let the outcomes of the matching process be such that ι players are matched with μ players playing action m. We now show that $\overline{u}^{\iota}_{t+1} > \overline{u}^{\mu}_{t+1}$:
First, suppose $N^{\iota}_{t} < \frac{N}{2} - N^{\iota}_{t}$. Then $\overline{u}^{\iota}_{t+1} = u(m',m) > \overline{u}^{\mu}_{t+1}$, from the first inequality of lemma 3.5. Second, suppose $N^{\iota}_{t} \geq \frac{N}{2} - N^{\iota}_{t}$. Then

$$\overline{u}^{\iota}_{t+1} = \frac{(\frac{N}{2} - N^{\iota}_{t})u(m',m) + (2N^{\iota}_{t} - \frac{N}{2})u(m',m')}{N^{\iota}_{t}} > \frac{(\frac{N}{2} - N^{\iota}_{t})u(m,m') + \frac{N}{2}u(m',m')}{N^{\mu}_{t}} = \overline{u}^{\mu}_{t+1},$$

where the inequality follows from lemma 3.5 and from the fact that $N^{\iota}_{t} < \frac{N}{2}$.

Let all the μ players playing action m, be offered the opportunity to revise their type. They will change type and become ι players playing action m'. Hence, the state in period $t+2$, is $s(t+2) = \{0,...,s^{\iota}_{m'},s^{\mu}_{m'},...,0\}$, with $s^{\iota}_{m'} + s^{\mu}_{m'} = N$.

We have now shown that when $m' = \max\{\overline{m}|\overline{m} \in br(s^{-n}(t))\} < m$, the system converges to a monomorphic state with $s_{m'} = N$. If $m' \in \mathcal{M}^{*}$, i.e. m' is a pure action Nash equilibrium. If not, let all myopic optimizers learn again. They now all face a state $s^{-n}(t+2) \prec s^{-n}(t)$, since $m' < m$. Hence, for each myopic optimizer n, $\max br\left(s^{-n}(t+2)\right) \leq \min br\left(s^{-n}(t)\right) \leq \max br\left(s^{-n}(t)\right) = m' < m$. Furthermore, since m' is not a Nash equilibrium, $\max br\left(s^{-n}(t+2)\right) \neq m'$, leading to $\min br\left(s^{-n}(t+2)\right) < m'$. By repeating the procedures given above, and using

the fact that the state space is finite, we conclude that the system ends up in a monomorphic state which is a pure strategy Nash equilibrium. This ends part 1 of the proof.

Part 2: Suppose the state is in any polymorphic state at time t. Let all imitators learn. With positive probability they all switch to the same action m'. From now, do not administer any learning draws to these imitators. They thus stay with their action m'. In consecutive periods, as long as not all myopic optimizers are taking a best reply against the current population state, let only one myopic optimizer update at a time.[8] In these periods, let no players get the opportunity to revise their type. From KR's Proposition 8 and Theorem 2, we know that the system converges to a state where all myopic optimizers are playing the same action, say action m, which is a best reply to the polymorphic state $s(t') = \{0, ..., s^{\iota}_{m'} = N^{\iota}, \ldots, s^{\mu}_{m} = N^{\mu}, ..., 0\}$. Since $m \in br(s(t'))$, there exist outcomes of the matching process such that $\overline{u}^{\mu} \geq \overline{u}^{\iota}$. Let such a match be realized. In the next period, let all imitators learn. They will switch to the action with highest average payoff. Hence, with positive probability all imitators switch to m and the monomorphic state $s\left(t'+1\right) = \{0, ..., s^{\iota}_{m}, s^{\mu}_{m}, ..., 0\}$, with $s^{\iota}_{m} + s^{\mu}_{m} = N$ is reached. □

4. Selection in the Presence of Mutations

We now focus on the model with mutations, i.e. we fix $\delta > 0$ and set $\varepsilon > 0$. Then we characterize the limit invariant distribution $\phi^{*}_{\delta} = \lim_{\varepsilon \to 0} \phi_{\delta}\left(\varepsilon\right)$ of the Markov chain $P\left(\varepsilon, \delta\right)$ specified by the model. First we show that when a population consisting of only imitators play a strict supermodular game the equilibrium selection procedure proposed by KMRY has almost no predictive power at all. This result should be contrasted with our main results, where we no longer restrict the model, in the sense that we let both imitators and myopic optimizers be present. In this case we obtain a clear-cut selection result: the payoff dominant equilibrium is selected as the unique long run equilibrium, and convergence is fast.

Before we proceed, we focus attention on the restriction of payoffs in the stage game. We assume that no off-diagonal payoff is large enough (in absolute value) to be able to upset an equilibrium with only one mutation. In a large population setting, this is equivalent to only restricting the off-diagonal payoff to the highest action. The supermodular structure of the stage game then ensures that the other off-diagonal payoffs are within the required range. So,

[8] Since the myopic optimizers are taking best responses to the (possible mixed) population state, the outcomes of the matching process are irrelevant for these players' calculation of best replies.

we assume that

$$\overline{u}^M = \frac{(N-2)u(M,M) + u(M,m')}{N-1} \geq u(m',M) \text{ for all } m' \neq M.$$

This in fact restricts $u(M,m')$ to be above the lowerbound $(N-1)\,u(m',M) - (N-2)u(M,M)$, which is guaranteed when we assume that $u(M,m') \geq \max_{\widetilde{m}} (N-1)\,u(\widetilde{m},M) - (N-2)u(M,M)$.

Note that when N is large the above expression simplifies to $\overline{u}^M \simeq u(M,M) \geq u(m',M)$, which is satisfied since M is a (strict) Nash equilibrium. Furthermore note that supermodularity alone does not restrict all payoff pairs in the stage game and permits that e.g. $u(M,1)$ becomes small enough to upset the equilibrium (M,M) with only one mutant who plays action 1. As a last point, note that even in cases where the assumption is violated, we can e.g. raise the payoff $u(M,M)$ to a level at which again the assumption is met, without changing the set of pure action NE or the Pareto ranking of the elements in this set.

Proposition 4.1. *Let only imitators be present in the population at all times. Then, there exists some $\overline{N} > 0$, such that for all $N > \overline{N}$, the payoff dominant equilibrium is one of the long run equilibrium states and convergence is of order ε^2, for small enough $\varepsilon > 0$. Furthermore, the set of long run equilibria contains at most $m-1$ monomorphic states.*

Proof.

The proof has two parts. First, we show that two mutations suffice to reach the payoff dominant equilibrium from any other state. Second, we show that two mutations are enough to upset the payoff dominant equilibrium. Note that since we are in a large population setting, one mutation is not enough to upset any strict Nash equilibrium.

Part 1: From Proposition 3.1 we know that from any state $s \in S$, the unperturbed dynamics will bring the system to a monomorphic state. In a monomorphic state, $\overline{u}_m(t) = u(m,m)$ for $m \in \{1,2,...,M\}$. Furthermore, since $u(M,M) > u(m,m)$ for $m \in \{1,2,...,M-1\}$, two mutations to M, which are matched realize the highest average payoff. Finally, with positive probability all players get the learning draw in the next period and switch to playing action M. Hence, the state $s = \{0,...,0,N\}$ is reached with two mutations.

Part 2: We now show that 2 mutations are enough to upset the payoff dominant equilibrium. Suppose the state at time t is $s(t) = \{0,\ldots,0,N\}$, and two players mutate to strategy m' and m'' respectively, where $u(m',m'') > u(M,M)$. Generically, such payoffs may be present in the stage game. Now suppose further that these two mutant players are matched. Again, let all players receive learning draw. Since $\overline{u}_{m'}(t) = u(m',m'') > \overline{u}_M(t) = u(M,M)$, all players switch

action. Now, it might be that $\overline{u}_{m'}(t) > \overline{u}_{m''}(t) = u(m'', m')$, leading to all players adopting action m' in period $t+1$, or it might be that $\overline{u}_{m'}(t) < \overline{u}_{m''}(t)$, leading to all players adopting action m'' at time $t+1$ (of course there is also the non-generic possibility that $\overline{u}_{m'}(t) = \overline{u}_{m''}(t)$, in which case with positive probability all players switch to m' or m''). Thus, we can reach any monomorphic state $s = \{0, ..., 0, s^{\iota}_{m'} = N, 0, \ldots, 0\}$ (or $s = \{0, ..., 0, s^{\iota}_{m''} = N, 0, \ldots, 0\}$) from $s = \{0, ..., 0, N\}$ with two mutations, as long as there is an entry $u(m', m'') > u(M, M)$, $m' \neq M$, (and of course $m'' \neq M$ and $m' \neq m''$, which has to hold since (M, M) is the payoff dominant Nash equilibrium in the stage game) somewhere in the payoff matrix of the stage game. Once the system is in a monomorphic state $s = \{0, ..., 0, s^{\iota}_{m'} = N, 0, \ldots, 0\}$, we can repeat the above argument. This way, with two mutations, the system can reach monomorphic states $s = \{0, ..., 0, s^{\iota}_{\widetilde{m}'} = N, 0, \ldots, 0\}$, when a payoff pair $u(\widetilde{m}', \widetilde{m}'') > u(m', m')$, $\widetilde{m}' \neq m'$, is present in the stage game. Note that it is possible that $u(\widetilde{m}', \widetilde{m}'') < u(M, M)$, i.e. with two mutations, the system can reach certain monomorphic states it could not reach directly from $s = \{0, ..., 0, N\}$ via other 'intermediate' monomorphic states $s = \{0, ..., 0, s^{\iota}_{m'} = N, 0, \ldots, 0\}$. Although the move trough an 'intermediate' monomorphic state requires a total of four mutations, it still only requires two simultaneous mutations at a time, which makes it an order ε^2 event for small $\varepsilon > 0$, just as switching directly to a state $s = \{0, ..., 0, s^{\iota}_{m'} = N, 0, \ldots, 0\}$.

In general strict supermodular games, this argument leads to many (almost all) monomorphic states being reachable with only two simultaneous mutations being necessary at a time. Also, strict supermodular stage games exists, in which the selection result is 'stronger' in the sense that less monomorphic states can be reached, since for these monomorphic states $s = \{0, ..., 0, s^{\iota}_{m} = N, 0, \ldots, 0\}$ it holds that $u(m, m') < u(m'', m'')$, for any $m'', m' \in \mathcal{M}$, $m' \neq m''$. Obviously, such a state cannot be reached through having two simultaneous mutations occur. Note that this is always the case for $m = 1$, since $(1, 1)$ is a Nash equilibrium, which guarantees that $u(\widetilde{m}, 1) < u(1, 1) < u(m', m')$ for any $m' > 1$. Thus, we have that the set of long run equilibria when only imitators are present in the population, contains at most $m - 1$ monomorphic states and no other states. Depending on the exact specification of the stage game, the set of long run equilibria may contain strictly less than $m - 1$ states.

Furthermore, it follows directly from the above arguments that the order of convergence is ε^2. □

Thus, we see that selection through a mutation process in a setting where only imitators are present does not lead to any clear-cut equilibrium predictions, as was the case in the absence of

mutations. Note that a similar result would hold in a setting with both imitators and myopic optimizers being present, when myopic optimizers would have a flat mutation rate equal to that of the imitators. In such a setting, a procedure similar to the one described in part 2 of the proof above, would lead to the selection of a strict subset of the set of symmetric pure action Nash equilibria of the stage game.

These results are contrasted by our main proposition below, which does lead to a clear-cut selection of a single equilibrium.

Proposition 4.2. *Consider a strict supermodular game satisfying Assumption A-D. Then, there exists some* $\overline{N} > 0$, *such that for all* $N > \overline{N}$ *the payoff dominant equilibrium is selected as the unique long run equilibrium and the expected wait is of an order less than* ε^{-2}, *for all small enough* ε.

Proof.

The structure of the proof is as follows. First, we show that two mutants are enough to get the system to the payoff dominant (PD) equilibrium. Then we argue that for a large subclass of supermodular games, more than two mutants are needed in order to get the system out of this equilibrium. However, there also exist strict supermodular games for which two mutants can get the system out of the PD equilibrium. As a last point we show that even for these games the probability of moving out of the PD equilibrium is arbitrarily much smaller than that of moving towards the PD equilibrium, when the probability of mutations goes to zero in the limit.

Consider a small probability of mutation, $\varepsilon > 0$. We have from Proposition 3.4 that the system moves to a monomorphic state costlessly, i.e. without any mutations being involved, and stays there as long as no mutations occur. Consider the system in such a monomorphic state labelled $all - m$, not being a state in which all players play action M. Note that in a monomorphic state myopic optimizers (imitators) can costlessly invade a population consisting solely of imitators (myopic optimizers) and most of the time the population fractions of myopic optimizers and imitators will both be approximately $\frac{1}{2}$. Suppose at time t two mutants occur. They both play M, meet and realize a payoff above the average payoff of all other players. Now, let an even number of imitators get the learning draw on the action level at time $t+1$. They will switch to action M. Also let an even number of myopic optimizers get the learning draw at time $t+1$. They may or may not update their action, depending on what is a best reply to the (now) mixed population state. Let all of these myopic optimizers update to the same action m'. Thus, at time $t+1$, the number of M-playing players has increased compared to time t,

both from imitators updating to action M and possibly from myopic optimizers updating to action M. Since we still have an even number of M-playing individuals, let each M-playing player be matched to another M-playing player. The same line of reasoning as above again leads to (the possibility of) an increase of the number of players playing action M from time $t+1$ to time $t+2$. Furthermore, with positive probability all myopic optimizers that play m' are matched to another m'-playing myopic optimizer. In case $m' \neq M$, $\overline{u}_t^{\mu} < \overline{u}_t^{\iota}$ at $t+2$ and then let some myopic optimizers receive a learning draw on the type level. They will switch type and become imitators and will choose to play either action m or action M, since both of these actions are still played by imitators[9]. If $m' = M$, it depends on the numbers $s_M^{\mu}(t+1)$ and $s_M^{\iota}(t+1)$ whether $\overline{u}_t^{\mu} \lesseqgtr \overline{u}_t^{\iota}$ (or equality), and thus it depends on these numbers whether myopic optimizers that get the learning draw on the type level will change type. However, in this case a best reply to the population state will remain to play M, because of the monotonicity of the best responses. Thus with positive probability, the number of M-playing players strictly increases along this path and with positive probability the increase in the number of M-playing individuals continues until all players play M. The system has now reached a population state in which in each match the PD equilibrium is played. Note that in this state there can be either only imitators or both imitators and myopic optimizers present in the population. Furthermore, note that the transition to the PD equilibrium along this path only involved two mutations and that it does not matter which players have mutated. Since in a monomorphic state myopic optimizers (imitators) can costlessly invade a population consisting solely of imitators (myopic optimizers), when the probability of mutation goes to zero, the most likely two players to mutate are two myopic optimizers, since they mutate at a rate $\varepsilon^{\beta(m)} > \varepsilon$.

If there are no out-of equilibrium payoff in the stage game which are higher than the payoff in the PD equilibrium, more than two mutants are needed in order to lead the system to a different equilibrium. In case there are out-of-equilibrium payoffs which are higher than the PD equilibrium payoff, at least two mutants are needed. We now show that there exist strict supermodular games for which exactly two mutants are needed. W.l.o.g. label m and m' such that $u(m,m') > u(M,M)$ and $u(m,m') > u(m',m)$ and let, at time t, one player mutate to action $m \neq M$ and another one to action $m' \neq M$, $m' \neq m$. Furthermore, let these two players be matched. This leads to $\overline{u}_m(t) > \overline{u}_M(t)$ and $\overline{u}_m(t) > \overline{u}_{m'}(t)$. Consequently at time $t+1$, the imitators that get the learning draw will update to action m. With positive probability enough imitators in the population get a learning draw to shift the best reply (to the new population

[9]Of course there is the possibility that no imitators playing action m are present any longer, in which case all type-switching myopic optimizers start to play action M.

state at time $t+2$) to be to play m. Then myopic optimizers will also update to action m as soon as they get a learning draw on the action level. Furtheron this path, we let the m-players be matched among themselves and let one of the M players be matched to the m'-player, who has not yet received a learning draw and therefore still plays action m'. Then, there exist strict supermodular stage games for which the payoffs are such that still $\overline{u}_m(t) > \overline{u}_M(t)$ (these are games for which e.g. $u(M,m') \ll u(M,M)$). Continuing along such a path, with positive probability leads to the equilibrium in which all players play action m. Thus there are strict supermodular games for which the PD equilibrium can be upset by just two mutations. For these games, we provide exact calculations showing that the PD equilibrium is harder to upset than other equilibria.

We now determine the probability that two mutations take place. Label the random variable X as the number of mutants among the imitators and Y as the number of mutants among the myopic optimizers. Thus

$$\Pr\left(X_t\left(\varepsilon\right)=k\right)=\binom{N_t^\iota}{k}\varepsilon^k\left(1-\varepsilon\right)^{N_t^\mu-k},\ k=0,1,\ldots,N_t^\iota$$

and

$$\Pr\left(Y_t\left(\widetilde{\pi}_t^i,\ i=1,\ldots,N_t^\mu\right)=k\right)=\sum_{i=1}^{N_t^\mu}Z_t\left(\widetilde{\pi}_t^i\right),$$

where

$$Z_t\left(\widetilde{\pi}_t^i\right)=\begin{cases}1, & \text{with prob } \varepsilon^{\beta\left(\widetilde{\pi}_t^i\right)},\\ 0, & \text{with prob } 1-\varepsilon^{\beta\left(\widetilde{\pi}_t^i\right)},\end{cases}$$

and where $\widetilde{\pi}_t^i$ is the rank number of the payoff π_t^i at time t of the myopic optimizers labelled as player i. Note that Y is an addition of independent Bernoulli trials with different parameters. Therefore, $Y\left(\cdot\right)$ is a function of the (rank number of) the payoffs to all myopic optimizers. Thus, when all myopic optimizers get the same payoff $\widetilde{\pi}_t^i$, and consequently have the same $\beta\left(\widetilde{\pi}_t^i\right)$, Y has a binomial distribution (just like X), with parameters $\beta\left(\widetilde{\pi}_t^i\right)$ and N_t^μ. Therefore, at any equilibrium state, $Y\left(\cdot\right)$ is a binomial random variable. The higher the equilibrium payoff, the lower the $p=\varepsilon^{\beta(.)}$ parameter of this binomial variable.

Two mutations arise with probability

$$\Pr(X+Y=2)=\Pr\left(X=2,Y=0\right)+\Pr\left(X=1,Y=1\right)+\Pr\left(X=0,Y=2\right).$$

At the PD equilibrium, all myopic optimizers play the PD equilibrium action and, as such, they receive the highest possible equilibrium payoff in the stage game. Therefore, their $\beta\left(\widetilde{\pi}\right)$ has reached it's maximum among equilibrium payoffs. Label this maximum $\beta^{\max}=\max$

$\{\beta(\widetilde{\pi})\,|\widetilde{\pi}$ is an equilibrium payoff$\} < 1$. This causes the (implicit) mutation rate among myopic optimizers to be at its minimum among equilibrium payoffs, namely $\varepsilon^{\beta^{\max}}$. The probability of having a myopic optimizer mutate is thus of a lower order in the PD equilibrium than it is in any other equilibrium. Now, consider what happens when we take ε to zero in the limit.

At the PD equilibrium $\varepsilon^{\beta^{\max}} > \varepsilon$ and thus all myopic optimizers have a higher mutation rate than all imitators (Note that the birth & death process sees to it that the system can always costlessly move to a state with $N_t^{\mu} > 2$). Therefore, $\Pr(X = 2, Y = 0)$ and $\Pr(X = 1, Y = 1)$ go to zero at a higher rate than $\Pr(X = 0, Y = 2)$, and we only have to focus on this last probability. The same is true for a move towards the PD equilibrium. Now, note that this probability is of a lower order when we look at a move away from the PD equilibrium than it is when we look at a move towards the PD equilibrium, since at the PD equilibrium (per definition) the equilibrium payoff is higher. Thus we have shown that

$$\frac{\Pr(X = 0, Y = 2|s(t) \text{ is the PD eq})}{\Pr(X = 0, Y = 2|s(t) \text{ is another eq})} = \frac{O\left(\varepsilon^{\beta^{\max}}\right)}{O\left(\varepsilon^{\beta'}\right)} = O\left(\varepsilon^{2\beta^{\max}-2\beta'}\right) \to 0, \text{ as } \varepsilon \to 0,$$

since $0 < \beta' < \beta^{\max} < 1$ and thus $0 < 2\beta^{\max} - 2\beta' < 2$. This means that the PD equilibrium is the unique long run equilibrium. Furthermore, from the above considerations, (see also Ellison (1993)), it follows directly that the expected wait until this equilibrium is reached is of an order strictly less than ε^{-2}. □

5. Discussion

Ideally, what one would like to obtain in an evolutionary model with noise is a clear-cut equilibrium selection result for a general class of games combined with fast convergence, (to make the equilibrium prediction plausible). Our main results, stated in Proposition 4.2 have such a flavor: when the two simple behavior rules, imitation and myopic optimization, coexist in the population and the latter type experiments more in low payoffs states, the payoff dominant equilibrium is selected as the unique long run equilibrium for the class of strict supermodular games. Furthermore, the expected waiting time until the payoff dominant equilibrium is reached is relatively short, even in the limit as the population size grows large. To understand the mechanisms underlying these results, Proposition 4.2 should be contrasted with the equilibrium selection and convergence results obtained in a population consisting of only myopic optimizers (Kandori and Rob (1995)) and with only imitators (Proposition 4.1). In the former, the geometry of the best response correspondence helps to identify the long run equilibrium (which may or may not be the payoff dominant one), and the convergence time increases with

the population size. Therefore, the waiting time until the long run equilibrium is reached is relatively long. In the latter, the interaction of evolutionary forces and mutations has almost no selection power at all, but convergence to the set of long run equilibria is fast even when the population size grows large. Imitation is therefore the force that drives the convergence result: imitators react to payoff differences and amplify the power by which mutations may upset *any* equilibrium configuration. Myopic optimizers, on the other hand, increase the equilibrium prediction in the class of strict supermodular games by ensuring that the limit set of the unperturbed dynamics is in one-to-one correspondence with the set of pure (symmetric) Nash equilibria in the stage game. Finally, the specific mutation model used in this paper, further improves the equilibrium prediction by ensuring that when the mutation rate approaches zero, the payoff dominant equilibrium is selected as the unique long run equilibrium.

References

Bergin, J. and B.L. Lipman (1996). Evolution with state-dependent mutations. *Econometrica 64*, 943–956.

van Damme, E. and J. Weibull (1998). Evolution with mutations driven by control costs. Discussion paper, nr. 9894, CentER, Tilburg University.

Ellison, G. (1993). Learning, local interaction, and coordination. *Econometrica 61*, 1047–1071.

Freidlin, M. and A. Wentzell (1984). *Random Perturbations of Dynamical Systems*. New York: Springer Verlag.

Gale, D. and R.W. Rosenthal (1999). Experimentation, imitation, and stochastic stability. *Journal of Economic Theory 84*, 1–40.

Harsanyi, J.C. and R. Selten (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: The MIT Press.

Kandori, M., G.J. Mailath, and R. Rob (1993). Learning, mutation and long run equilibria in games. *Econometrica 61*, 29–56.

Kandori, M. and R. Rob (1995). Evolution of equilibria in the long run: A general theory an applications. *Journal of Economic Theory 65*, 383–414.

Milgrom, P. and J. Roberts (1990). Rationalizability, learning, and equilibrium in games with strategic complementaries. *Econometrica 58*, 1255–1277.

Robles, J. (1998). Evolution with changing mutation rates. *Journal of Economic Theory 79*, 207–223.

Robson, A.J. and F. Vega-Redondo (1996). Efficient equilibrium selection in evolutionary games with random matching. *Journal of Economic Theory 70*, 65–92.

Topkis, D.M. (1979). Equilibrium points in nonzero-sum n-person submodular games. *SIAM Journal of Control and Optimization 17*, 773–787.

Vega-Redondo, F. (1995). Expectation, drift, and volatility in evolutionary games. *Games and Economic Behavior 11*, 391–412.

Vega-Redondo, F. (1996). *Evolution, Games, and Economic Behavior*. New York: Oxford University Press.

Vives, X. (1990). Nash equilibrium with strategic complementarities. *Journal of Mathematical Economics 19*, 305–321.

Young, H.P. (1993). The evolution of conventions. *Econometrica 61*, 57–84.