Implementation, Elimination of Weakly Dominated Strategies and Evolutionary Dynamics

Antonio Cabrales[¤]
Universitat Pompeu Fabra and ELSE

Giovanni Ponti[y]
University College London and ELSE

December 18, 1996

Abstract

This paper is concerned with the realism of mechanisms that implement social choice functions in the traditional sense. Will agents actually play the equilibrium assumed by the analysis? As an example, we study convergence and stability properties of Sjöström's (1994) mechanism, on the assumption that boundedly rational players find their way to equilibrium using monotonic learning dynamics. This mechanism implements most social choice functions in economic enviroments using as a solution concept the iterated elimination of weakly dominated strategies (only one round of deletion of weakly dominated strategies is needed). There are, however, many sets of Nash equilibria whose payoffs may be very different from those desired by the social choice function. We show that many equilibria in all the sets of equilibria we describe are the limit points of trajectories that have completely mixed initial coniditions. The initial conditions that lead to these equilibria need not be very close to the limiting point. Furthermore, even if the dynamics converge to the "right" set of equilibria, it still can converge to quite a poor outcome in welfare terms.

[¤] This paper was written while Cabrales was visiting the Centre for Economic Learning and Social Evolution at the UCL. He would like to thank for the hospitality and encouragement. The financial support of the EU HCM program is gratefully acknowledged.

[y] The financial support of the Commission of the European Communities, under the project "Human Capital Mobility - Fellowship ERBCHBICT941575" is gratefully aknowledged.

1 Introduction

The theory of implementation studies the problem of designing decentralized institutions through which certain socially desirable objectives can be achieved. These social arrangements should be able to operate without extensive knowledge by the principal about the agents, and in a variety of environments. The principal should ensure that the rules of the game are respected by the agents, and such rules should be designed so that it is in the best interest of the agents to take actions that lead to the socially desirable outcome, given the environment.

More precisely, a social choice rule is implemented by a mechanism (game-form) if for every possible environment (preference pro¯le) the solution (set of equilibrium outcomes) of the mechanism coincides with the set of outcomes of the social choice rule for every possible environment.

One of the problems that are faced by the implementation theorist is the choice of a solution concept. This is no trivial matter because for di®erent solution concepts the range of social choice rules that can be implemented varies dramatically. As Moore (1990) points out \choice rules are unlikely to be implementable in dominant strategy equilibrium if the domain is very rich and/or the choice rule is e±cient". In the case of single-valued choice rules he also notes that \the move from dominant strategy to Nash may not help at all: only the restricted class of strategy-proof choice may be Nash implementable". If the solution concept is more re¯ned, then the domain of the social choice rule can be much larger. In fact, the social choice rule domains are considerably enlarged for subgame-perfect implementation (Moore and Repullo 1988), and even more so when the solution concept is the iterative deletion of weakly dominated strategies (Abreu and Matsushima 1994, Jackson et al.1994, Sjöström 1994). In fact, as Sjöström (1994) says: \With enough ingenuity the planner can implement `anything"'.

The question that arises then is whether the equilibrium concept chosen is a good one for the game in object. One way to answer the question is to assume that agents are boundedly rational and that they adjust their actions over time through some trial and error learning procedure. One can then analyze under which conditions the actions that lead to the socially desirable outcomes are played asymptotically, if at all. Research in implementation theory has paid little attention to the problem of how an equilibrium is reached. Some exceptions are the papers of Muench and Walker (1984) and de Trenqualye (1988), who study the local stability of the Groves and Ledyard (1977) mechanism, and Cabrales (1996), who studies the global convergence of the canonical mechanism (Maskin 1977, Repullo 1987) of Nash implementation and the mechanisms of Abreu and Matsushima (1992, 1994).

In this paper we study the convergence and stability properties of the Sjöström's (1994) mechanism when one assumes that the players are boundedly rational and the dynamics are monotonic (Samuelson and Zhang 1992, Weibull 1995). One particularly well known

member of the family of monotonic dynamics is the replicator dynamics of evolutionary game theory (Taylor and Jonker, 1978). These dynamics have been given a learning theoretic foundation by Börgers and Sarin (1993), and they can also be interpreted as a model of imitation (Schlag, 1994). Sjöström's (1994) mechanism and the one that Jackson et al. (1994) study for separable environments are very similar and most of our results would generalize easily for that mechanism as well.

We concentrate on Sjöström's mechanism for several reasons. One is that the conditions for implementation are quite weak. Although the environments that are permitted are not universal, they are rich enough for most economic problems. Furthermore, this reduction in the domain permits the author to implement the social choice rule with a "bounded" game and thus makes it immune to the criticisms of Jackson (1992). Finally, although the solution concept is the iterated elimination of (weakly) dominated strategies (it also implements in undominated Nash equilibria), it only needs one round of deletion of weakly dominated strategies (the first). This last feature of the mechanism makes it particularly attractive since under some assumptions of imperfect knowledge of agents (either because of payoff uncertainty as in Dekel and Fudenberg, 1990, or through lack of perfect common knowledge of rationality as in Börgers, 1994) the appropriate solution concept implies one round of deletion of weakly dominated strategies and then the iterated deletion of strictly dominated strategies.

In Sjöström's (1994) mechanism the agents are arranged to announce their preferences and those of their two closest neighbors. The mechanism is designed in such a way that a truthful report of one's own preferences is weakly dominant (it does not affect one's payoff, except in a set of states which is called totally inconsistent, and in those states it is preferable to report them truthfully). Since in this mechanism it is advantageous to report the same preferences about your neighbors that they are reporting about themselves it is clear that the only equilibrium that survives the first round of deletion of weakly dominated strategies is the truth-telling one. There are, however, many other Nash equilibria. For every preference profile R, there is a component (i.e. a closed and connected set) of equilibria in which all agents report the preferences for their neighbors indicated in R and they report the preferences about themselves indicated in R with high enough (this need not be very high) probability. The reason for this is that the mechanism makes it important that all agents match their neighbors' announcements about themselves, but the report about oneself is only important in some unlikely (totally inconsistent) state.

We show that many equilibria in all the components of equilibria we have described are the limit points of trajectories of the learning dynamics that have completely mixed initial conditions (that is, initial conditions that give strictly positive weights to all possible messages). Although the general results are local, we can show by example (the game in Figure 1, Sjöström, 1994) that the initial conditions that lead to these equilibria need not be close to the limiting point. Furthermore, and perhaps more worrying, the equilibria which belong to the same component as the completely truthful report are not outcome equivalent to such equilibrium, as they yield payoffs that are significantly different (lower)

to the payo®s of the social choice functions outcome. Therefore, even if the dynamics converge to the \right" component of equilibria, it still can converge to quite a poor outcome in welfare terms.

One could naively expect that evolution would eliminate weakly dominated strategies. The reason why this doesn't happen is that the weakly dominant strategy grows faster than the dominated one only if the totally inconsistent states are met often enough by the players. But the weight of the totally inconsistent states is also decreasing over time since people are learning to avoid such states. It may be that they decrease fast enough so that the push towards the weakly dominant strategy is not enough to make the dominated strategy disappear.

The fact that evolution does not eliminate weakly dominated strategies has been known since at least Nachbar (1990). Samuelson (1993) discusses the issue of elimination of weakly dominated strategies in evolutionary games. Binmore et al. (1995) have shown the implications of these ¯ndings for the ultimatum bargaining game. In particular, they provide a numerical example, based on the classic \chain store game", in which a) there are trajectories of the replicator dynamics which converge to the Nash equilibrium component in which the players choose a weakly dominated strategy with positive probability and b) in the presence of mutations, such component may even exhibit asymptotic stability properties. These results are more than a theoretical curiosity. Binmore and Samuelson (1996) note that: \the experimental evidence is now strong that one cannot rely on predictions that depend on deleting weakly dominated strategies".

The remainder of the paper is arranged as follows. In section 2 we introduce some notation, we describe the mechanism and we make the assumptions about the dynamics. In section 3 we fully characterize (for all interior initial conditions) the set of limit points of the dynamics for the game in Figure 1, Sjöström 1994, to be considered a simpli¯ed version of the mechanism. In section 4 we give local results (for some interior initial conditions) for the set of limit points of the dynamics for the general game. In section 5 we describe the asymptotic stability properties of the sets of limit points in the presence of mutations. Finally, section 6 concludes, together with an appendix containing the proofs of the relevant propositions.

2 The model and the dynamics

We will ¯rst introduce some notation, the assumptions on preferences and the game form proposed by Sjöström (1994). Then we introduce the assumptions we make on the dynamics.

The only important change we make in the presentation with respect to Sjöström (1994) is that we make the assumptions on a Von Neumann-Morgenstern utility function instead of on the preference relation. We do this because we need to specify the payo® functions

for mixed strategies, since the dynamics are de¯ned on the mixed strategy simplex.

There is a set $I \acute{} f1;:::;ng$ (with $n \cdot 3$) of agents and a set $A \in <^m_+$ of feasible consumption plans. The preferences of agent $i \in I$ are represented with a (Von Neumann-Morgenstern) utility function $v_i : A \pounds \copyright_i \to <$, where $\copyright_i$ speci¯es a ¯nite set of possible utility functions. An element R_i of $\copyright_i$ is meant to represent the preferences of agent i over A. A preference pro¯le is a vector $R = (R_1;:::;R_n)$, where $R_i \in \copyright_i$. The preference pro¯les will be common knowledge among the agents. The following assumptions are made concerning preferences and feasible consumption pro¯les.

Assumption p.1. The set of feasible consumption pro¯les is convex. For all $a; a' \in A$ and for all $\lambda \in [0;1]$ then $\lambda a + (1 - \lambda)a' \in A$:

Assumption p.2. The preferences represented by $R_i \in \copyright_i$ are complete and transitive.

Assumption p.3. The preferences represented by $R_i \in \copyright_i$ are strictly convex. That is, for any $a; a' \in R^m_+$ and for all $\lambda \in (0;1)$ if $a \neq a'$ and $v_i(a;R_i) \geq v_i(a';R_i)$ then $v_i(\lambda a + (1 - \lambda)a';R_i) > v_i(a';R_i)$:

Assumption p.4. For any $R_i \in \copyright_i$ if $a \geq 0$ and $a \neq 0$ then $v_i(a;R_i) > v_i(0;R_i)$:

Assumption p.5. Preference reversal. For any $R_i; R'_i \in \copyright_i$ if $R_i \neq R'_i$ then there are $a; a \in A$ such that $v_i(a;R_i) > v_i(a;R_i)$ and $v_i(a;R'_i) > v_i(a;R'_i)$:

For any set $B \mu <^m_+$ and any $R_i \in \copyright_i$ a choice correspondence is de¯ned as follows: $c(B;R_i) \acute{} fa \in Bj$ for all $b \in B; v_i(a;R_i) \geq v_i(b;R_i)g$:

A social choice function is a mapping $f : \copyright \to A$; where $f(R) \acute{} (f_1(R);::::;f_n(R))$. A mechanism is a pair $\gamma \acute{} (M;\Re)$, where $M \acute{} \pounds_{i2I} M_i$ and $\Re : M \to A$; where $\Re(m) \acute{} (\Re_1(m);::::;\Re_n(m))$. M_i is the message space of agent i and $\Re$ is the outcome function. A mechanism and a preference pro¯le de¯ne a game.

Let $M_{\neg i} \acute{} M_1 \pounds ::: \pounds M_{i-1} \pounds M_{i+1} \pounds ::: \pounds M_n$. Given a mechanism γ and a preference pro¯le R, we say that m_i is weakly dominated for some set of messages $F \acute{} \pounds_{i2I} F_i \mu M$ if there exists a message $m'_i \in F_i$ such that $v_i(\Re_i(m'_i;m_{\neg i});R_i) \geq v_i(\Re_i(m_i;m_{\neg i});R_i)$ for all $m_{\neg i} \in F_{\neg i}$ and there is some $m^{\alpha}_{\neg i} \in F_{\neg i}$ such that $v_i(\Re_i(m'_i;m^{\alpha}_{\neg i});R_i) > v_i(\Re_i(m_i;m^{\alpha}_{\neg i});R_i)$: De¯ne the set $U_i(F : (\gamma;R)) \acute{} fm_i \in F_i jm_i$ is not weakly dominated in F for the game $(\gamma;R)g$:

The message m_i is a best response for player i, to $m_{\neg i} \in M_{\neg i}$ if $v_i(\Re_i(m_i;m_{\neg i});R_i) \geq v_i(\Re_i(m'_i;m_{\neg i});R_i)$ for all $m'_i \in M_i$. A message pro¯le m is a Nash equilibrium (NE) if m_i is a best response to $m_{\neg i}$ for all $i \in N$. A message pro¯le $m \in M$ is an undominated Nash equilibrium (UNE) for the game $(\gamma;R)$ if it is a Nash equilibrium and $m_i \in U_i(M : (\gamma;R))$: Let $UNE(\gamma;R) \acute{} f\Re(m) \in Ajm$ is an UNE for the game $(\gamma;R)g$:

We say that a mechanism $(M;\Re)$ implements a social choice function f in undominated Nash equilibrium if for all $R \in \copyright$, $f(R) = UNE(R)$.

For the iterated deletion of weakly dominated strategies let $U_i^1(\cdot;R) = U_i(M:(\cdot;R))$, and if $U_i^k(\cdot;R)$ has been defined for $k \geq 1$, let $U_i^{k+1}(\cdot;R) = U_i(\times_{j\neq i} U_j^k(\cdot;R):(\cdot;R))$. Let $U_i^1(\cdot;R) = \cap_{k=1}^{1} U_i^k(\cdot;R)$. Let $IWD(\cdot;R) = \{\alpha(m) \in A | m_i \in U_i^1(\cdot;R) \text{ for all } i\}$.

We say that a mechanism $(M;g)$ implements a social choice function f with iterated deletion of weakly dominated strategies if for all $R \in \mathcal{C}$, $f(R) = IWD(R)$.

We now construct a mechanism.

Let $M_i = \mathcal{C}_{i-1} \times \mathcal{C}_i \times \mathcal{C}_{i+1}$, so that each individual announces the preferences of her two neighbors, and let members of M_i and M be denoted m_i and m respectively. A generic strategy is therefore $m_i = (R_{i-1}^i; R_i^i; R_{i+1}^i)$. A K-tuple of messages $\{m_{j_1}; \ldots; m_{j_K}\}$ is totally consistent if whenever agents $i,k \in \{j_1; \ldots; j_K\}$ both announce the preference of player $j \in I$, then $R_j^i = R_j^k$. On the other hand, a K-tuple of messages $\{m_{j_1}; \ldots; m_{j_K}\}$ is totally inconsistent if whenever agents $i,k \in \{j_1; \ldots; j_K\}$ both announce the preference of player $j \in I$, then $R_j^i \neq R_j^k$.

Consider $R_i, R_i' \in \mathcal{C}_i$, where $R_i \neq R_i'$. By assumption p.6 there are $a, \bar{a} \in A$ such that $v_i(a;R_i) > v_i(\bar{a};R_i)$ and $v_i(a;R_i') > v_i(\bar{a};R_i')$. We can choose a and $\bar{a}$ so that $v_i(a;R_i) > v_i(a';R_i)$ for all a' in the line segment between a and $\bar{a}$. Given this pair $(a,\bar{a})$ let $\bar{\gamma}_i(R_i,R_i') = \{b \in \prec_m^+ | b = \lambda a + (1-\lambda)\bar{a}, \text{ for } \lambda \in [0,1]\}$. By construction, for all $R_i, R_i' \in \mathcal{C}_i$, $c(\bar{\gamma}_i(R_i,R_i');R_i) \neq c(\bar{\gamma}_i(R_i,R_i');R_i')$. Let $\acute{A}(i;m) \equiv (R_1^n; R_2^1; \ldots; R_{i-1}^i; R_{i+1}^i; R_{i+2}^{i+1}; \ldots; R_n^{n-1})$ and for every i and m_{i-1}, define

$$B_i(m_{i-1}) = \begin{cases} f_i(\acute{A}(i;m)) & \text{if } m_{i-1} \text{ is totally consistent} \\ \bar{\gamma}_i(R_i^{i-1}; R_i^{i+1}) & \text{if } m_{i-1} \text{ is totally inconsistent} \\ \frac{1}{n} f_i(\acute{A}(i;m)) & \text{Otherwise} \end{cases}$$

Now we can define α:

$$\alpha_i(m) = \begin{cases} c(B_i(m_{i-1}); R_i^i) & \text{if } R_{i-1}^i = R_{i-1}^{i-2} \text{ and } R_{i+1}^i; R_{i+1}^{i+1} \\ 0 & \text{otherwise} \end{cases}$$

To understand the mechanism notice that the only time when the choice of an announcement R_i^i has any effect on payoffs is when m_{i-1} is totally inconsistent. In that case, the outcome is the optimal choice within the set $\bar{\gamma}_i(R_i^{i-1}; R_i^{i+1}))$ according to the announced R_i^i. For this reason announcing the true preference R_i^* can never hurt. Furthermore, for every alternative announcement $R_i^i = \hat{R}_i$, there is some totally inconsistent m_{i-1} with $R_i^{i-1} = R_i^*$ and $R_i^{i+1} = \hat{R}_i$ and the set $\bar{\gamma}_i(\cdot;\cdot)$ is constructed in such a way that $c(\bar{\gamma}_i(R_i^*;\hat{R}_i);R_i^*)$ is strictly preferred to $c(\bar{\gamma}_i(R_i^*;\hat{R}_i);\hat{R}_i)$. Therefore, a message $m_i = (R_{i-1}^i; \hat{R}_i; R_{i+1}^i)$ is weakly dominated by a message $m_i = (R_{i-1}^i; R_i^*; R_{i+1}^i)$, that is, untruthful announcements about oneself are weakly dominated.

Once these weakly dominated strategies are eliminated and all agents announce the true preferences about themselves, $R_j^j = R_j^*$, it is strictly dominated to announce untruthful preferences about the neighbors, $R_{i+1}^i \neq R_{i+1}^* = R_{i+1}^{i+1}$ or $R_{i-1}^i \neq R_{i-1}^* = R_{i-1}^{i-1}$, since disagreeing with the neighbors is punished with the 0 consumption bundle.

These two facts establish the main theorem in Sjöström (1994).

Proposition 0. Let f be an arbitrary social choice function. The mechanism described above implements f in UNE and in IWD.

It is important to notice, for the discussion we will undertake below, that the set of states for which not announcing the true preferences about oneself is weakly dominated are themselves states that typically produce very bad outcomes for the opponents (at least one of them will have 0 consumption, and probably many). If agents learn to avoid totally inconsistent states very fast, there is no incentive to tell the truth about oneself. The mechanism we have described puts a lot of emphasis in consensus announcements, since disagreement is punished with 0 consumption, and truth-telling is only rewarded in a set of states which need not be very prominent in the minds of the players. That is precisely the reason why convergence to outcomes of the social choice function may fail to occur. This conflict is typical of other mechanisms that implement in the iterated deletion of weakly dominated strategies, like Abreu and Matsushima (1994).

We now move on to the characterization of the evolutionary dynamics we analyze.

Fix a given preference profile $R \in \prec$: Let $x_i^{m_i}$ be the probability assigned by agent i to message m_i; and $x_i \in \Delta_i$ be a mixed strategy for agent i (where Δ_i denotes the $|M_i| - 1$-dimensional simplex which describes player i's mixed strategy space). Let also $x_{-i} \in \times_{j \neq i} \Delta_j$ be a mixed strategy profile for agents other than i; with $x \equiv (x_i; x_{-i}) \in \Delta \equiv \times_{i \in I} \Delta_i$. Finally, let $u_i(x_i; x_{-i}) = \sum_{m \in M} v_i(\theta_i(m_i; m_{-i}); R_i) \prod_{j \in I} x_j^{m_i}$:

We formalize player i's behavior in terms of the mixed strategy $x_i(t)$ she adopts at each point in time. The vector $x(t)$ will then describe the state of the system at time t, defined over the state space Δ, with Δ^0 denoting the relative interior of Δ, that is, the set of completely mixed strategy profiles. We make the following assumption:

Assumption d.1 The evolution of $x(t)$ is given by a system of continuous-time differential equations:

$$\dot{x} = D(x(t)) \tag{1}$$

We require that the autonomous system (1) satisfies the standard regularity condition, i.e., D must be i) Lipschitz continuous with ii) $\sum_{k \in S_i} D_i^k(\cdot) = 0$ and iii) $\lim_{x^{m_i} \to 0} \frac{\dot{x}^{m_i}}{x^{m_i}}$ well-defined and finite.[1] Furthermore, D must also satisfy the following requirements:

Assumption d.2. D is a regular (payoff) monotonic selection dynamic. More explicitly, let $g_i(m_i; x_{-i}(t)) = \frac{\dot{x}_i^{m_i}(t)}{x_i^{m_i}(t)}$. Then for all m_i, m_i', we have that

$$\text{sign}\left[g_i(m_i; x_{-i}(t)) - g_i(m_i'; x_{-i}(t))\right] = \text{sign}\left[u_i(m_i; x_{-i}(t)) - u_i(m_i'; x_{-i}(t))\right]$$

[1] A useful implication of this regularity assumption is that the solution of the dynamical system leaves Δ, as well as Δ^0, invariant (and, a fortiori, forward invariant): any solution path starting from Δ (Δ^0) does not leave Δ (Δ^0). This property will prove to be useful to obtain some of the results of the paper.

Assumption d.2 is a common assumption in the evolutionary literature and we will not dwell on it (see, for example, Samuelson and Zhang 1992 and Weibull, 1995).

Assumption d.3. Let $Y_i(m_i; m_i^0) = \{x_{-i} | u_i(m_i; x_{-i}(t)) - u_i(m_i^0; x_{-i}(t)) = 0\}$. Then, for all $\pm > 1$:

$$\frac{\lim_{d(x_{-i}(t); Y_i(m_i; m_i^0)) \to 0} \sup [g_i(m_i; x_{-i}(t)) - g_i(m_i^0; x_{-i}(t))]}{\text{sign}[u_i(m_i^0; x_{-i}(t)) - u_i(m_i; x_{-i}(t))] (\ln |u_i(m_i^0; x_{-i}(t)) - u_i(m_i; x_{-i}(t))|)^{\pm}} > -1$$

Assumption d.3 is less standard in the evolutionary literature and we will expand on it when we discuss Proposition 4 because it will be helpful to understand why weakly dominated strategies need not disappear in the limit. What assumption d.3 says is that if the difference in payoffs between two strategies is going to zero a rate $\exp[-n]$, the difference in growth rates has to go to zero at least at a rate $1/n^{\pm}$. Continuity and assumption d.2 demand that strategies that have the same payoff grow at the same rate, but they impose no requirements on the speed at which the difference in growth rates goes to zero as the difference in payoff go to zero. Assumption d.3 can be satisfied even if the sensitivity of growth rates to payoffs is much higher than linear around zero (as would be implied, for example, by the replicator dynamics and other aggregate monotonic dynamics).

Assumption d.4. $x(0) \in \Phi^0$

Finally, Assumption d.4, which is also standard in the evolutionary literature, is necessary because it excludes the possibility that the selection dynamic acts only on a subset of the strategy space. This possibility arises because the system is forward invariant, and therefore a strategy that has zero weight at time zero would also have zero weight at all subsequent times. We want to avoid this possibility because the selection dynamics would be operating on a game which might be qualitatively different from the game we are trying to analyze.

3 An example.

We prefix the analysis of the dynamics of the mechanism with the following example, taken from Sjöström (1994), p. 504, which is intended to convey the essence of our results. There is a unit of good which has to be divided among three players: 1, 2 and 3. The (Von-Neumann Morgenstern) utility functions of players 1 and 2 are linear in the amount of good consumed, and this is common knowledge among the players and the planner. The utility function of player 3 may have one of two possible types, either linear on the amount of good consumed (we index these preferences by the number 1) or linear until the amount of good consumed is 1/3, for consumptions larger than 1/3 the utility remains constant at the value 1/3, since the agent becomes satiated (the index for

these preferences is 0). The true preferences of player 3 are common knowledge among the players, but the planner does not know them.

The social choice function recommends the consumption vector (1=3; 1=3; 1=3) for preferences of type 1 and (1=4; 1=4; 1=2) for preferences of type 0. Notice that this social choice function is such that agent 3 would like to conceal her preferences, and therefore the planner needs a nontrivial mechanism to elicit the true preferences.

The mechanism proposed by Sjöström requires the three players to make a simultaneous statement about the preferences of player 3. Let $m_i^1; i \in I$ represent the message in which preferences of type 1 are announced, with m_i^0 denoting the announcement of type 0 preferences. Figure 1 illustrates the outcome function. We will assume for the analysis that the true preferences are of type 1 and therefore Figure 1 is also the payoff function of a game, which we call Γ:

Figure 1

Sjöström's Example: game Γ:

Player 1 picks a row, player 2 a column, and player 3 picks a matrix. We first notice that the mechanism leads to a game which is weakly dominance solvable, in the sense that it can be reduced to a single cell, corresponding to the truth-telling equilibrium outcome, by the iterated deletion of weakly dominated strategies. Unlike other weakly solvable games, this procedure yields, in this example, a unique outcome, independently of the order of removal of strategies. Player 3 first deletes her (weakly) dominated strategy m_3^0 (the other agents have no dominated strategies at this stage). Once m_3^0 is removed, the strategies m_1^0 and m_2^0 for players 1 and 2 become strictly dominated. The unique strategy profile selected in this way is $(m_1^1; m_2^1; m_3^1)$. Notice, however, that the strategy profile $(m_1^0; m_2^0; m_3^0)$ is also an equilibrium, and that this equilibrium yields a higher payoff for agent 3 than $(m_1^1; m_2^1; m_3^1)$.

Given that each player has only two strategies in her support, with an abuse of notation we set $x_i \equiv x_i^{m_i^1}$.[2] We first characterize the set of Nash equilibria of the game:

Proposition 1. The set NE of Nash equilibria of Γ is the union of precisely two disjoint components NE^0 and NE^1, where:

$$NE^0 \equiv \{x \in \Delta \mid x_1 = x_2 = 0; x_3 \leq \tfrac{3}{7}\}$$
$$NE^1 \equiv \{x \in \Delta \mid x_1 = x_2 = 1; x_3 \leq \tfrac{1}{2}\}$$

Proof. See the Appendix. □

[2] The fact that each player has only two available options will also allow us to express the dynamics in terms of the payoff difference between player i's truthful and untruthful strategy, which we call $\Delta_i(x(t))$ (i.e. $\Delta_i(x(t)) \equiv u_i(m_i^1, x_{-i}(t)) - u_i(m_i^0, x_{-i}(t))$).

Denote with $RE(\dot{\iota})$ the set of restpoints of $\dot{\iota}$ under any monotonic dynamic. It is straightforward to show that $RE(\dot{\iota})$ contains (together with all the pure strategy pro-¯les) only the following components: $RE^0 \acute{} fx \in \Delta | x_1 = x_2 = 0; x_3 \in [0;1]g$ and $RE^1 \acute{} fx \in \Delta | x_1 = x_2 = 1; x_3 \in [0;1]g$: Our task is to study the asymptotics of a monotonic selection dynamic whose initial state lies in the relative interior of the state space:

Proposition 2. Any solution $x(t; x(0))$ of a monotonic selection dynamics $\underline{x} = D(x)$ with completely mixed initial conditions converges asymptotically to NE.

Proof. See the Appendix2

If initial conditions are completely mixed, we then know that the evolutionary dynamics will eventually converge to a Nash equilibrium of the game. In the following section we extend the result to the more general setting of Sjöström's mechanism.

4 Local results for the general game

In this section we show that the results of the previous section generalize locally. Proposition 3 characterizes some components of Nash equilibria for the game induced by the mechanism in Sjöström (1994), which we described in section 2. Any message pro¯le in which the agents are unanimous in the (arbitrary) preference pro¯le they announce, R^{α}, (or, more appropriately, the preferences they announce about their neighbors and themselves are taken from the pro¯le R^{α}) is an equilibrium. Furthermore, a mixed strategy pro¯le in which every agent mixes between messages consistent with R^{α} and other preference pro¯les that only di®er in the announcement they make about their own preferences is also an equilibrium, as long as R^{α} is given a high enough weight. As we showed in the example, the weight given to R^{α} need not be very high. The equilibria in a component are not payo® equivalent, since disagreeing with a neighbor (an event with nonzero probability in the mixed strategy equilibria) results in a punishment. Proposition 4 shows that any of the previous equilibria that gives enough weight to R^{α} is the limit point of some interior path for a monotonic selection dynamic. Figure 2 shows that the initial condition need not be very close to the limit point.

Let $\hat{R}$ be the true preference pro¯le and R^{α} an arbitrary preference pro¯le.
Let $m_i^{\alpha} = (R_{i-1}^{\alpha}; R_i^{\alpha}; R_{i+1}^{\alpha})$; $U_i = \max_R v_i(f_i(R); \hat{R}_i)$ and $U_{in} = \max_R v_i\left(\frac{1}{n} f_i(R); \hat{R}_i\right)$. m_i^{α} is a consensus announcement by agent i, U^i is utility associated to the most preferred outcome from the social choice function for agent i with true preferences $\hat{R}^i$ and U^{in} is utility associated to the most preferred consumption bundle among those that result from dividing the bundles assigned by the social choice function by n.

$$S_i = fm_i \in M_i | R_i^{i-1} = R_{i-1}^{\alpha}; R_i^{i+1} = R_{i-1}^{\alpha}g$$

and $S_i^1 = fm_i \in M_i | m_i \notin S_i g$ The set S_i is the set of all mixed strategies in which

announcements about the neighbors agrees with R^{α}, and S_i^1 is the complement of S_i with respect to M_i.

$$S_i^{k_i} = \{x_i | x_i^{m_i} = 0, \text{ for all } m_i \notin S_i \text{ and } x_i^{m_i^{\alpha}} > k_i\},$$

where we assume

$$(k_i)^n > \frac{U_{jn} - u_j(0; \hat{R}_j)}{v_j(f_i(\hat{A}(j; R^{\alpha})); \hat{R}_j) - v_j(0; \hat{R}_j) + U_{jn} - v_j(0; \hat{R}_j)}$$

for all i and all $j \neq i$. The set $S_i^{k_i}$ is the set of all mixed strategies in which announcements about the neighbors agrees with R^{α}, and the probability of announcing R_i^{α} is higher than k_i.

Proposition 3. For all $\hat{R}, R^{\alpha} \in <$ and $x_i \in S_i^{k_i}$, x is a Nash equilibrium of $(\Gamma, \hat{R})$.

Proof. See the Appendix.2

Now we prove that not only are there other Nash equilibria, but that elements in those components can be reached by paths starting in the interior to the simplex. By assumption d.2 we know that for all $h_v > 0$ with $u_i(m_i; x_{-i}(t)) - u_i(m_i^0; x_{-i}(t)) < -h_v$, there is $h_g > 0$, such that $g_i(m_i; x_{-i}(t)) - g_i(m_i^0; x_{-i}(t)) < -h_g$.

Let h_v be a constant such that $0 < h_v < \min_{i,R} v_i(f_i(\hat{A}(i; R^{\alpha})); \hat{R}_i) - v_i(0; \hat{R}_i)$. Let the corresponding h_g and

$$H = \max_i \left\{ \left(\frac{U_i - v_i(0; \hat{R}_i) + h_v}{v_i(f_i(\hat{A}(i; R^{\alpha})); \hat{R}_i) + U_i - 2v_i(0; \hat{R}_i)} \right)^{\frac{1}{n-1}} \right\}.$$

Note that by definition $H \in [0, 1]$.

By Assumption d.3 we know that there exists $\mu_i(m_i; m_i^0) > 0$ such that if $|u_i(m_i; x_{-i}(t)) - u_i(m_i^0; x_{-i}(t))| < \mu_i(m_i; m_i^0)$

$$\frac{\lim_{d(x_{-i}(t); Y_i(m_i; m_i^0)) \to 0} \sup |g_i(m_i; x_{-i}(t)) - g_i(m_i^0; x_{-i}(t))|}{|\ln |u_i(m_i^0; x_{-i}(t)) - u_i(m_i; x_{-i}(t))||^{\pm}} > 1$$

Let $0 < \mu < \min_{i; m_i; m_i^0} \mu_i(m_i; m_i^0)$.

For any set $\mathbb{C}_i \in M_i$, let $x_{\mathbb{C}_i} = \sum_{m_i \in \mathbb{C}_i} x_i^{m_i}$, and $L = \exp\left[\frac{h_g(\pm - 1) - \ln\left(U_i + \sum_{j \neq i} x_j^{S_j}(0) \frac{H}{x_i^{m_i^{\alpha}}(0)} \right)}{\pm - 1} \right]$

Proposition 4. Assume that for all i, $x_i^{m_i^{\alpha}}(0)$ is big enough so that, $x_i^{m_i^{\alpha}}(0) L > H$ and $U_i + \sum_{j \neq i} x_j^{S_j}(0) \frac{H}{x_i^{m_i^{\alpha}}(0)} < \mu$. Under these conditions we have that

a) For all $m_i \in \hat{S}_i$, $\frac{x_i^{m_i}(t)}{x_i^{m_i}(0)} < \exp[-h_g t]\frac{H}{x_i^{m_i^*}(0)}$ for all t and all i

b) $x_i^{m_i^*}(t) > H$ for all t

c) $\frac{x_i^{m_i^*}(t)}{x_i^{m_i}(t)} < \frac{x_i^{m_i^*}(0)}{x_i^{m_i}(0)} L$ for all t and all $m_i \in S_i$

Proof. See the Appendix.2

Part a) of the Proposition says that the weight of any strategy in $\hat{S}_i$ decreases over time at a rate higher than h_g. This is important because the strategies for which not telling the truth about oneself is dominated are all in $\hat{S}_j$, so if the weight of these strategies decrease over time, the payoff advantage of the dominating strategy disappears over time, and makes it possible for a dominated strategy to have positive limiting weight.

Part b) ensures that the weight of m_i^* is always high enough. If the weight of m_i^* is high enough, then the strategies in the sets $\hat{S}_j$ have a lower payoff than strategies in S_j since an announcement about your neighbor that does not coincide with her announcement about herself is punished.

In fact parts a) and b) reinforce each other. While m_i^* keeps having a high enough weight, the weight of strategies in $\hat{S}_j$ decreases, and if strategies in $\hat{S}_j$ decrease fast enough the weight of m_i^* does not go below a certain bound. All of this provided that m_i^* started with high enough weight, which as Figure 2 demonstrates, need not be very high.

Notice that part b) guarantees that pure strategy equilibria in the "wrong" component are attractors of interior paths. Part c) says that the weight of m_i^* in the limit is less than 1, and therefore some mixed strategy equilibria are attractors as well if the initial conditions give sufficiently little weight to strategies in $\hat{S}_j$. This guarantees that even if there is convergence to the "right" component it need not be to the pure strategy equilibrium, and remember that the equilibria are not payoff equivalent (the mixed strategy equilibria have lower expected payoff because agents are punished for announcing discordant preferences).

The convergence to the mixed equilibria can happen because the payoffs to all strategies in S_i are similar if the weight of strategies in $\hat{S}_j$ is small, and by a) the weight of strategies in $\hat{S}_j$ is decreasing. So even though m_i^* has a payoff advantage, the advantage goes to zero over time, and assumption d.3 guarantee that it does not accumulate fast enough.

If d.3 didn't hold, equilibria which do not implement the social choice function may fail to be a limit point for the dynamics. Convergence to the "wrong" equilibrium obtains only if a weakly dominated strategy for player i (call it m, and call m' the strategy that weakly dominates m) gets positive weight in the limit. But along the way to the limit the strategy against which m and m' differ (call it y) has also positive weight (since the system is regular, and therefore forward invariant). So by assumption d.2 the growth rate of m' is larger than the growth rate of m. The weight of m'=m is the integral of the difference in growth rates of m' and m. If the limiting value of this integral is infinite the limiting value of $x_i^m(t)$ would be zero. But the weight of y (and thus the difference in

payo®s) may be going to zero, thus the weight of m^0/m is an integral of a function that goes to zero, which may be ¯nite.

Assumption d.3 describes how the growth rates have to relate to payo®s (when di®erences in payo®s are small) so that the limiting value of a dominated strategy is not zero. Assumption d.3 would hold, for example, if the growth rates were linear in the payo®s, as it happens with the replicator dynamics. But the requirement is much weaker than that, because it is only a local requirement around zero, and because the rate at which growth rates go to zero with payo®s can be much higher than linear. In other words, even if the growth rates were much more sensitive to payo® di®erences (around zero) than the replicator dynamics allow, assumption d.3 could still be satis¯ed.

The elimination of a weakly dominated strategy in an evolutionary context requires that the strategy against which the dominated strategy gives a lower payo® than the dominating strategy has to appear su±ciently often or that its appearance has to provoke a dramatic enough reduction in the dominated strategy.

5 More on the example (stability with/out drift).

In the previous section, we have extended the convergence result contained in Proposition 2, and we have shown that the limit points of the dynamics for interior initial conditions are generally di®erent from the outcomes intended by the planner. We now go back to the example in order to test the stability properties of NE. To do so, some further terminology is needed:

Definition 1. Let C be a closed set of restpoints in ¢ of the dynamics (1): Then: (i) C is called (interior) stable if, for every neighborhood O of C, there is another neighborhood U of C, with U ½ O;such that the trajectory of any initial point in U \ ¢ (U \ $¢^0$) remain inside O;

(ii) C is called (interior) attracting if is contained in an open set O such that every initial point in O and also in (the interior of) ¢ evolves under D to C. C is globally (interior) attracting if every point in (the interior of) ¢ evolves under D to C;

(iii) C is called (interior) asymptotically stable if it is (interior) attracting and (interior) stable.

To simplify the analysis, we set additional conditions on the dynamics, which is the purpose of the following assumption, (which replaces assumptions d.1-5):

Assumption d.5. The evolution of x(t) is given by the following system of continuous-time di®erential equations:

$$\dot{x}_i \equiv D_i(x(t); \lambda) = x_i(t)\,(1 - x_i(t))\,[\phi_i(\cdot) + \lambda\,(\bar{\alpha}_i - x_i(t))] \qquad (2)$$

with $\xi \geq 0$; $\beta_1 = \beta_2 = \frac{1}{2}$ and $\beta_3 = \beta \in (0,1)$.

In words: the evolutionary dynamic is now composed of two additive terms. The first represents the standard replicator dynamic, while the second term ensures that, at each point in time, each strategy is played with positive probability, no matter how it performs against the current opponents' mixed strategy profile (i.e. it points the dynamic toward the relative interior of the state space Σ). Following Binmore and Samuelson (1996), this latter term is called drift: it opens the model to the possibility of a heterogeneity of behaviors. Binmore et al. (1995), derive an analogous system in the following way. At each point in time, a fixed proportion of players (of measure $\frac{\xi}{1+\xi}$) is replaced by new individuals whose aggregate behavior is represented by a generic, constant, completely mixed strategy (i.e. β_i), while the rest of the population aggregate behavior follows the replicator dynamics. The relative importance of the drift is measured by ξ; which we refer to as the drift level. We assume ξ to be "very small", reflecting the fact that all the major forces which govern the dynamics should be captured by the evolutionary dynamic defined by D; which here takes the form of the replicator dynamics.

We check how the model reacts to the introduction of such a perturbation. The stability analysis of the replicator dynamics with drift will give us information about the effects of small changes in the vector field on the equilibria of the system defined by the replicator dynamic (in other words, it will test the structural stability of such equilibria). To simplify the exposition, β_1 and β_2 have been chosen to be 1/2 since only the value of β_3 turns out to be genuinely significant.

We start our analysis on the stability properties of NE looking at the case of the replicator dynamic without drift (i.e. when $\xi = 0$): We know from Proposition 2, that NE is globally interior attracting, since it attracts every interior path under any monotonic selection dynamic (of which the replicator dynamic is a special case). We now take a closer look at the stability properties of each component of Nash equilibria separately (i.e. NE^0 and NE^1) :

Figure 2

The replicator dynamic and game γ

Figure 2 shows a phase diagram describing trajectories of the replicator dynamic starting from some interior initial conditions. The Nash equilibrium component NE^0 (NE^1) is represented by a bold segment in the bottom-left (top-right) corner of the state space Σ: First notice that, as we know from Proposition 2, all trajectories converge to a Nash equilibrium of the game. Moreover, the diagram shows (consistently with Proposition 4) that there are some trajectories of the replicator dynamic which converge to NE^0; the Nash equilibrium component in which both players 1 and 2 deliver the false message with probability 1. However, this latter component is not asymptotically stable, as can be easily spotted from the diagram. Trajectories starting arbitrarily close to NE^0; provided

$x_3 > \frac{3}{7}$; will eventually converge to the truth-telling component. We summarize the key properties of these trajectories in the following proposition:

Proposition 5. Under the replicator dynamic (i) NE^1 is interior asymptotically stable, whereas (ii) NE^0 is not.

Proof. See the Appendix. 2

We now move to the analysis of the replicator dynamic with drift:

Figure 3

The dynamic with drift and game ¡

Let ¯ 2 (0;1) be a generic element of the space of the feasible perturbations. Figure 3 shows the trajectories of the replicator dynamic with drift under two di®erent speci¯cations of ¯: Diagram 3b) represents a situation in which, in the proximity of NE^0; the drift against m_i^0 is uniform across players, where in diagram 3a) the drift against m_3^0 is lower. As the ¯gures show, there is a local attractor close to NE^1 in both cases. Moreover, none of the elements of NE^0 is a restpoint of the dynamic with drift in ¯gure 3b), while ¯gure 3a) shows that there is an additional local attractor which belongs to NE^0 : trajectories starting close to NE^0 converge to it, as it happens in the case of the replicator dynamics without drift.

We are interested in the convergence and stability properties of (2) when ¸ ! 0; considering two di®erent con¯gurations of the drift parameter ¯ :

$$\text{CASE A}: \bar{\beta} \in \left[0; \frac{23 - 4\sqrt{30}}{49}\right)$$
$$\text{CASE B}: \bar{\beta} \in \left(\frac{23 - 4\sqrt{30}}{49}; 1\right]$$

Given $\frac{23 - 4\sqrt{30}}{49} = .0222673$, CASE A depicts a situation in which, for small values of x_i; the drift against the untruth-telling strategy is substantially lower for player 3 than for her opponents.

4In the following proposition we characterize the set of restpoints of the dynamic with drift, together with their stability properties:

Proposition 6. Let $\hat{RE}(\bar{\beta})$ be the set of restpoints of (2) for ¸ su±ciently close to 0. The following properties hold:

a)$\forall \bar{\beta} \in (0;1)$; $\hat{RE}(\bar{\beta})$ contains an element of NE^1; which is also asymptotically stable.

b) under CASE A $\hat{RE}(\bar{\beta})$ contains also two additional restpoints, both belonging to NE^0; one of which is asymptotically stable.

Proof. See the Appendix:2.

There is a striking similarity between the content of Proposition 6 and the ¯ndings of Binmore et al. (1995), as we pointed out in the introduction. They analyze the chain store game, in one of whose equilibrium components a player selects a weakly dominated strategy with positive probability. This component is interior attracting. Moreover, like our NE^0; such component fails to be interior asymptotically stable, but for certain parameter values it may be asymptotically stable when the system is slightly perturbed. Given the failure of asymptotic stability without perturbations, one would expect any perturbation to move the system away from the unstable component and the weakly dominated strategy to become extinct. Proposition 6 tells us that evolutionary game theory does not provide a ground for such claim. The intuition is similar to the one in Binmore et al. (1995). When there is drift, the strategies against which the weakly dominated strategy does poorly will have positive weight at all times and therefore the part of the dynamics that depend on payo®s pushes against the dominated strategy. But the drift may provide a direct push in favor of the dominated strategy (and more crucially, in favor of those strategies of the other players which do well against such dominated strategy). When the balance between these two forces is right, one gets a stable equilibrium with positive weight for the dominated strategy, as it happens in our example.

6 Conclusions

We have argued that there is room for doubt about the practicability of one the of the leading examples of implementation with iterated deletion of weakly dominated strategies when agents are boundedly rational. This result complements that obtained by Cabrales (1996) for the Abreu and Matsushima (1994) mechanism. Since Cabrales (1996) uses dynamics that are di®erent from those used here, it would be interesting to check if the results we obtain here extend to Abreu and Matsushima (1994) games. More generally, a deeper study with evolutionary tools of other mechanisms studied in the literature would enhance our understanding of the performance of these mechanisms with boundedly rational agents, a necessary step before mechanisms are used in real life.

Ideally one would like to design a game for which convergence to the preferred social outcome could be guaranteed for the learning protocols that agents use. To achieve this goal, it is necessary to conduct empirical and experimental studies that reveal how people adjust their play in games like that studied in detail in this paper. The history of actual social arrangements may also give clues as to how people learn in such environments. Different mechanisms for public good provision have existed for centuries in many countries. These considerations imply the need for a substantial program of future research.

References

D. Abreu and H. Matsushima (1992), \Virtual Implementation in Iteratively Undominated Strategies: Complete Information", Econometrica, 60, 993-1008.

D. Abreu and H. Matsushima (1994), \Exact Implementation", Journal of Economic Theory, 64, 1-19.

K. Binmore, J. Gale and L. Samuelson (1995), \Learning to Be Imperfect: the Ultimatum Game", Games and Economic Behavior, 8, 56-90.

K. Binmore and L. Samuelson (1996), \Evolutionary Drift and Equilibrium Selection", Institute for Advanced Studies, Vienna, Working Paper 26.

T. Börgers and R Sarin (1993), \Learning through Reinforcement and Replicator Dynamics", UCL Discussion Paper 93-13.

T. Börgers (1994), \Weak Dominance and Approximate Common Knowledge", Journal of Economic Theory, 64, 265-276.

A. Cabrales (1996), \Adaptive Dynamics and the Implementation Problem with Complete Information", Universitat Pompeu Fabra WP.

E. Dekel and D. Fudenberg (1990), \Rational Behavior with Payo® Uncertainty", Journal of Economic Theory, 52, 243-267.

T. Groves and J. Ledyard (1977), \Optimal Allocation of Public Goods: a Solution to the Free Rider Problem", Econometrica, 45, 783-809.

M. O. Jackson (1992), \Implementation in Undominated Strategies: A Look at Bounded Mechanisms", Review of Economic Studies, 59, 757-775.

M. O. Jackson, T. R. Palfrey and S. Srivastava (1994), \Undominated Nash Implementation in Bounded Mechanisms", Games and Economic Behavior, 6, 474-501.

Y. G. Kim and J. Sobel (1995), \An Evolutionary Approach to Pre-Play Communication", Econometrica, 63, 1181-1193.

E. Maskin (1977), \Nash Implementation and Welfare Optimality", mimeo, Massachusetts Institute of Technology.

J. Moore (1990), \Implementation in Environments with Complete Information", in J. J. La®ont ed., Advances in Economic Theory: Sixth World Congress, Econometric Society.

J. Moore and R. Repullo (1988), \Subgame Perfect Implementation", Econometrica, 58, 1083-1099.

T. Muench and M. Walker (1984), \Are Groves-Ledyard Equilibria Attainable?", Review of Economic Studies, 50, 393-396.

Nachbar, J. H. (1990), "Evolutionary Selection Dynamics in Games: Convergence and Limit Properties", International Journal of Game Theory, 19, 59-89.

R. Repullo (1987), \A Simple Proof of Maskin's Theorem on Nash Implementation", Social Choice and Welfare, 4, 39-41.

L. Samuelson (1993), \Does Evolution Eliminate Weakly Dominated Strategies?", in K. Binmore, A. Kirman and P. Tani (eds.) (1993) Frontiers of Game Theory, London, MIT Press.

L. Samuelson and J. Zhang (1992), \Evolutionary Stability in Asymmetric Games", Journal of Economic Theory, 57, 363-391.

K. Schlag (1994), "Why Imitate, and if so, How? Exploring a Model of Social Evolution", Friedrich-Wilhelms-Universitat Bonn, Discussion Paper No. B-296.

T. Sjöström (1994), \Implementation in Undominated Nash Equilibria without Integer Games", Games and Economic Behavior, 6, 502-511.

PD Taylor and LB Jonker (1978), `Evolutionary Stable Strategies and Game Dynamics", Mathematical Biosciences, 40, 145-156

P. de Trenqualye (1988), \Stability of the Groves and Ledyard Mechanism", Journal of Economic Theory, 46, 164-171.

J. Weibull (1995), Evolutionary Game Theory, Cambridge MA, MIT Press.

7 Appendix

Proof of Proposition 1. We already noticed that agent 3 has a weakly dominated strategy (namely, m_3^0). In particular, m_3^1 (truth-telling) makes agent 3 (strictly) better o® than m_3^0 (lying), unless agents 1 and 2 coordinate their actions completely, that is, unless they play m_i^0 $i = 1;2$ with probability 1 or they play m_i^1 $i = 1;2$ with probability 1, (in which case, 3 is completely indi®erent). This leads to the following lemma:

Lemma 1. No strategy pro¯le in which $x_3 \in (0;1)$ can be a Nash equilibrium unless $x_1 = x_2 = 1$ or $x_1 = x_2 = 0$, that is, unless agents 1 and 2 play the same strategy with probability 1.

With this consideration in mind, we construct the proof as follows: we ¯x the mixed strategy of player 3 and check what mixed strategies for player 1 and 2 can sustain a Nash equilibrium. Noting that

$$\Delta_1 \equiv u_1(m_1^1; x) - u_1(m_1^0; x) = \frac{1}{12}(x_2(x_3 - 1) + 7x_3 - 3) \tag{3}$$

$$\Delta_2 \equiv u_2(m_2^1; x) - u_2(m_2^0; x) = \frac{1}{12}(x_1(x_3 - 1) + 7x_3 - 3) \tag{4}$$

we can make the following observations:

a) When $x_3 < \frac{3}{7}$, m_i^0 (lying) yields a strictly higher payo® than m_i^1 for both 1 and 2, independently of what the other player does. Therefore, the strategy pro¯les in NE^0 (and only those) will be Nash equilibria.

b) When $x_3 = \frac{3}{7}$, m_1^0 yields a strictly higher payo® than m_1^1 unless $x_2 = 0$, and $x_2 = 0$ makes player 1 indi®erent between m_1^0 and m_1^1 (a symmetric argument holds for player 2). This excludes the possibility of $(1;1;\frac{3}{7})$ being a Nash equilibrium of the game, leaving $(0;0;\frac{3}{7}) \in NE^0$ as the unique Nash equilibrium when $x_3 = \frac{3}{7}$:

c) When $x_3 \in (\frac{3}{7};\frac{1}{2})$ there are no Nash equilibria. This happens because in this case if $x_1 = 1$, the best response of player 2 is $x_2 = 0$ and if $x_1 = 0$, the best response for player 2 is $x_2 = 1$: However, neither $(0;1;x_3)$ nor $(1;0;x_3)$ can be Nash equilibria when $x_3 \in (\frac{3}{7};\frac{1}{2})$ by Lemma 1.

d) $x_3 = \frac{1}{2}$: In analogy with the case where $x_3 = \frac{3}{7}$, it is an implication of Lemma 1 that $(1;1;\frac{1}{2}) \in NE^1$ is the unique Nash equilibrium when $x_3 = \frac{1}{2}$:

e) When $x_3 > \frac{1}{2}$ announcing m_i^1 (truth-telling) is optimal for $i = 1$ and 2, independently of what the other player does. Thus, the strategy pro¯les in NE^1 (and only those) will be Nash equilibria.

Since this exhausts all cases the result follows. 2

Proof of Proposition 2. To prove the proposition, it is enough to show that any interior trajectory converges. This is because, once convergence has been proved, we can apply the standard result "convergence implies Nash under any monotonic selection dynamics" (see, e.g. Weibull, 1995, Theorem 5.2 (iii)) to obtain the result.

We start by observing that the fact that the dynamic is forward invariant implies that $x_i(t)$ is always de¯ned and positive, for any nonnegative t. By monotonicity, $x_3(t)$ is also a positive, increasing function of t and bounded above by 1 (since m_3^1 is a weakly dominant strategy), therefore it must converge. This already implies convergence of player 3's mixed strategy. Let us denote $x_i^¤ ´ \lim_{t \to 1} x_i(t)$, when such a limit exists. Three alternative cases have to be discussed:

a) $x_3^¤ = 0$: If $x_3^¤ = 0$ there must be a time t^0 such that $x_3(t) < \frac{3}{7}$ for $t > t^0$. This implies that there is a $k > 0$ such that for all $t^0 > t$, $¢ ¦_i(x(t)) < ¡k$ for $i = 1; 2$. This implies, by monotonicity, $\lim_{t \to 1} x_i(t) = 0$ for $i = 1; 2$, thus $x^¤ = (0; 0; 0)$.

b) $x_3^¤ = 1$: By a similar argument, monotonicity implies $x^¤ = (1; 1; 1)$:

c) $x_3^¤ 2 (0; 1)$: We want to prove that $x_3^¤$ cannot converge to a value within this range unless the system converges to a Nash equilibrium. To do so (given the special features of our example) it is enough to show that, if $x_3^¤ 2 (0; 1)$ it then must be that both players 1 and 2 select, in the limit, the same pure strategy. Given that this result implies convergence of the entire mixed strategy pro¯le, the result follows. More formally, what we have to prove is contained in the following lemma:

Lemma 3. If $x_3^¤ 2 (0; 1)$ then:

either
$x_i^¤ = 0; i = 1; 2$ (CASE 0 hereafter)
or
$x_i^¤ = 1; i = 1; 2$: (CASE 1)

Proof. Let's assume, for the purpose of contradiction, that neither of the above statements is true. In that case, there must exist a sequence $ft_k g_{k=1}^1$ and a positive constant $² > 0$ such that either $x_i(t_k) > ²; i = 1; 2$ or $x_i(t_k) < 1 ¡ ²; i = 1; 2$ for all k (in other words, the system must stay in¯nitely often an ² away from the faces of ¢ in which player 1 and 2 play the same pure strategy). We already noticed that these are the only faces of ¢ in which both pure strategies for player 3 yield the same payo®: if the system stays away from them in¯nitely often along the solution path, it then must be that the cumulative payo® di®erence will grow unbounded as time goes to in¯nity. As we will see, this in turn implies (by monotonicity) that $x_3(t)$ will also reach, in the limit, its highest value, that is, $x_3^¤ = 1$; as a result of the extinction of the weakly dominated strategy m_3^0, which is a contradiction.

To show this, we ¯rst notice that the payo® di®erence $¢ ¦_i(x(t))$ is a continuous function of x(t) de¯ned over a compact set (¢): In the case of player 3, such function takes the

following form:

$$\phi_{\lambda_3}(x(t)) = \frac{(x_1(t) - x_2(t))^2 + x_1(t)(1 - x_1(t)) + x_2(t)(1 - x_2(t))}{6} \quad (5)$$

Take $g_M = \max_{i \in I; x_i \in \Phi_i}[|g_i(m_i; x_{-i}(t))|]$, i.e. the highest possible growth rate (in absolute value) over all strategies and players (we know a max exists, since also $g_i(\cdot)$ is continuous in Φ): Then define $\varepsilon_1; \varepsilon_2; \varepsilon_3$ and ε_4 as follows:

ε_1 solves $2\exp[-g_M \varepsilon_1] = \frac{\sigma^2}{2}$ (i.e. $\varepsilon_1 = \frac{\ln[2]}{g_M}$)

ε_2 solves $(1 - \sigma^2)\exp[-g_M \varepsilon_2] = \frac{\sigma^2}{2}$ (i.e. $\varepsilon_2 = \frac{\ln\left[-2+\frac{\sigma^2}{2}\right]}{g_M}$)

ε_3 solves $2\exp[g_M \varepsilon_3] = 1 - \frac{\sigma^2}{2}$ (i.e. $\varepsilon_3 = \frac{\ln\left[-\frac{1}{2}+\frac{1}{\sigma^2}\right]}{g_M}$)

ε_4 solves $(1 - \sigma^2)\exp[g_M \varepsilon_4] = 1 - \frac{\sigma^2}{2}$ (i.e. $\varepsilon_4 = \frac{\ln\left[\frac{2-\sigma^2}{2-2\sigma^2}\right]}{g_M}$)

and take $\Delta\varepsilon = \min[\varepsilon_1; \varepsilon_2; \varepsilon_3; \varepsilon_4]$; that is, set a lower bound for the time interval in which, after each t_k, $\frac{\sigma^2}{2} < x_i < 1 - \frac{\sigma^2}{2}; i = 1; 2$ and therefore $\phi_{\lambda_3}(x(t))$ still remains bounded away from 0 (i.e $\phi_{\lambda_3}(x(t)) > \frac{\sigma^2(1-\frac{\sigma^2}{2})}{3} > 0; \forall t \in [t_k; t_k + \Delta\varepsilon]$): Denote by $G_\sigma = \left\{x \in \Phi : \phi_{\lambda_3}(x) \geq \frac{\sigma^2(1-\frac{\sigma^2}{2})}{3}\right\}$. Now define:

$$\psi_i(x(t)) = \frac{\partial}{\partial t}\ln\left(\frac{x_i(t)}{1 - x_i(t)}\right) = \frac{\dot{x}_i(t)}{x_i(t)} - \frac{(1 - \dot{x}_i(t))}{1 - x_i(t)} = \frac{\dot{x}_i(t)}{x_i(t) - (x_i(t))^2}$$

i.e. the time derivative of the log of the ratio between the probabilities with which each of player i's pure strategies are played, which can be expressed in terms of the difference in the growth rates. Notice that also $\psi_3(x(t))$ will be a positive number bounded away from 0 infinitely often since, by assumption d.1, the difference in growth rates is a continuous function of x(t) defined on a compact set, which preserves the same sign of $\phi_{\lambda_3}(x(t))$: This implies that we can always define a constant $g_\sigma = \min_{x \in G_\sigma} \psi_3(x(t))$, with $g_\sigma > 0$ by assumption d.2. Also by assumption d.2 $\psi_3(x(t)) > g_\sigma \Leftrightarrow \phi_{\lambda_3}(x(t)) > \frac{\sigma^2(1-\frac{\sigma^2}{2})}{3}$: If we integrate the value of $\psi_3(x(t))$ over time we then obtain:

$$\lim_{t \to \infty} \int_0^t \psi_3(x(t))dt \geq \sum_{k=1}^{\infty} \int_{t_k}^{t_k + \Delta\varepsilon} \psi_3(x(t))dt > g_\sigma \sum_{k=1}^{\infty} \int_{t_k}^{t_k + \Delta\varepsilon} dt = \infty$$

which implies that $x_3^* = 1$; which leads to a contradiction.2

To summarize, Lemma 3 shows that, if $x_3^* \in (0; 1)$; $x_1(t)$ and $x_2(t)$ must converge (and therefore x(t) must converge to a Nash equilibrium). Since this exhausts all cases the result follows. 2

Proof of Proposition 3.

For all $\hat{x}_i$, such that $\hat{x}_i^{m_i} > 0$ only if $m_i \in S_i$ we have,

$$u_i(\hat{x}_i; x_{-i}) = u_i(x_i; x_{-i}) - \prod_{j \neq i} x_j^{m_j'} v_i(f_i(\hat{A}(i; R')); \hat{R}_i) + (1 - \prod_{j \neq i} x_j^{m_j'}) v_i(0; \hat{R}_i).$$

For all $\hat{x}_i$,

$$u_i(\hat{x}_i; x_{-i}) \leq (1 - \sum_{m_i \in S_i} \hat{x}_i) u_i(x_i; x_{-i}) + \sum_{m_i \in S_i} \hat{x}_i \left[\prod_{j \neq i} x_j^{m_j'} v_i(0; \hat{R}_i) + (1 - \prod_{j \neq i} x_j^{m_j}) U_{in} \right].$$

Then

$$u_i(\hat{x}_i; x_{-i}) - u_i(\hat{x}_i; x_{-i}) = \sum_{m_i \in S_i} \hat{x}_i \left[\prod_{j \neq i} x_j^{m_j'} (v_i(f_i(\hat{A}(i; R')); \hat{R}_i) - v_i(0; \hat{R}_i)) + (1 - \prod_{j \neq i} x_j^{m_j'})(v_i(0; \hat{R}_i) - U_{in}) \right]$$

which is greater than zero since by the definition of k_j,

$$\prod_{j \neq i} x_j^{m_j'} \geq \prod_{j \neq i} k_j \geq \frac{U_{in} - v_i(0; \hat{R}_i)}{v_i(f_i(\hat{A}(i; R')); \hat{R}_i) - v_i(0; \hat{R}_i) + U_{in} - v_i(0; \hat{R}_i)}. \qquad 2$$

Proof of Proposition 4. By contradiction.

Suppose that a) is the statement that stops being true earliest, that it does it for agent i and strategy $m_i \in \hat{S}_i$ and that the boundary time is t^0. Then it must be true that

$$\frac{x_i^{m_i}(t^0)}{x_i^{m_i}(0)} = \exp[-h_g t^0] \frac{H}{x_i^{m_i'}(0)}$$

Notice that for all t

$$u_i(x_i^{m_i}(t); x_{-i}(t)) - u_i(x_i^{m_i'}(t); x_{-i}(t)) \leq v_i(0; \hat{R}_i) \prod_{j \neq i} x_j^{m_j'}(t) + U_i(1 - \prod_{j \neq i} x_j^{m_j'}(t))$$

$$- (v_i(f_i(\hat{A}(i; R')); \hat{R}_i) \prod_{j \neq i} x_j^{m_j'}(t) + v_i(0; \hat{R}_i)(1 - \prod_{j \neq i} x_j^{m_j'}(t)))$$

$$= [U_i - v_i(0; \hat{R}_i) - \prod_{j \neq i} x_j^{m_j'}(t)(v_i(f_i(\hat{A}(i; R')); \hat{R}_i) + U_i - 2v_i(0; \hat{R}_i))]$$

But since b) is true for $t < t^0$

$$u_i(x_i^{m_i}(t); x_{-i}(t)) - u_i(x_i^{m_i'}(t); x_{-i}(t)) < \left[U_i - v_i(0; \hat{R}_i) - H^{n-1}(v_i(f_i(\hat{A}(i; R')); \hat{R}_i) + U_i - 2v_i(0; \hat{R}_i)) \right]$$

So we have that

$$u_i(x_i^{m_i}(t); x_{-i}(t)) - u_i(x_i^{m_i'}(t); x_{-i}(t)) < -h_v$$

which by assumption d.2 and the definition of h_v and h_g implies that

$$g_i(m_i; x_{-i}(t)) - g_i(m_i'; x_{-i}(t)) < -h_g.$$

which integrating from 0 to t^0 and given that $x_i^{m_i^*}(t^0) \cdot H$ implies that

$$\frac{x_i^{m_i}(t^0)}{x_i^{m_i}(0)} < \exp[\,\text{¡}\, h_g t^0]\frac{H}{x_i^{m_i^*}(0)}$$

This is a contradiction.

Suppose that b) is the statement that stops being true earliest, that it does it for agent i and that the boundary time is t^0. Then it must be true that $x_i^{m_i^*}(t^0) = H$.

First notice that for all $m_i \in S_i \setminus \{m_i^*\}$ since the payo®s of strategy m_i^* and other strategies in S_i di®er only when playing against strategies not in $S_{\text{¡}i}$

$$u_i(x_i^{m_i^*}(t); x_{\text{¡}i}(t)) \text{¡}\ u_i(x_i^{m_i}(t); x_{\text{¡}i}(t)) \text{¸}\ \text{¡}\, U_i \textstyle\prod_{j \neq i} x_j^{S_j}(t)$$

since a) holds for $t < t^0$

$$u_i(x_i^{m_i^*}(t); x_{\text{¡}i}(t)) \text{¡}\ u_i(x_i^{m_i}(t); x_{\text{¡}i}(t)) > \text{¡}\, U_i \left(\exp[\,\text{¡}\, h_g t]\frac{H}{x_i^{m_i^*}(0)} \textstyle\prod_{j \neq i} x_j^{S_j}(0)\right)$$

Since $U_i\left(\frac{H}{x_i^{m_i^*}(0)} \prod_{j \neq i} x_j^{S_j}(0)\right) < \mu$, this implies by assumption d.3 that

$$(g_i(m_i^*; x_{\text{¡}i}(t)) \text{¡}\ g_i(m_i; x_{\text{¡}i}(t))) > \text{¡}\left(\text{¡}\ \ln\left(U_i \frac{H}{x_i^{m_i^*}(0)} \textstyle\prod_{j \neq i} x_j^{S_j}(0)\right) + h_g t\right)$$

So by integration we have that

$$\frac{x_i^{m_i^*}(t^0)}{x_i^{m_i}(t^0)} \frac{x_i^{m_i}(0)}{x_i^{m_i^*}(0)} > \exp\left[\frac{1}{h_g(\pm \text{¡}\ 1)\ \text{¡}\ \ln\left(U_i \frac{H}{x_i^{m_i^*}(0)} \prod_{j \neq i} x_j^{S_j}(0)\right)}\right]^{\pm \text{¡}\ 1} = L$$

Adding over all strategies in S_i we have

$$\frac{x_i^{m_i^*}(t^0)}{x_i^{m_i^*}(0)} > \frac{x_i^{S_i}(t^0)}{x_i^{S_i}(0)} L = \frac{1 \text{¡}\ x_i^{S_i}(t^0)}{1 \text{¡}\ x_i^{S_i}(0)} L \text{¸}\ L$$

But this implies that $x_i^{m_i^*}(t^0) > H$ (using the assumption that $x_i^{m_i^*}(0)\ L > H$), which is a contradiction.

Suppose that c) is the statement that stops being true earliest, that it does it for agent i and that the boundary time is t^0. Then it must be true that $\frac{x_i^{m_i^*}(t^0)}{x_i^{m_i}(t^0)} = L \frac{x_i^{m_i^*}(0)}{x_i^{m_i}(0)}$

As before, notice that for all $m_i \in S_i \setminus \{m_i^*\}$ the payoffs of strategy m_i^* and m_i differ only when playing against strategies not in S_{-i}; so

$$u_i(x_i^{m_i^*}(t); x_{-i}(t)) - u_i(x_i^{m_i}(t); x_{-i}(t)) \geq U_i\left[\prod_{j \neq i} x_j^{S_j}(t)\right]$$

which by part a) of the proposition implies that for all $t < t'$

$$u_i(x_i^{m_i^*}(t); x_{-i}(t)) - u_i(x_i^{m_i}(t); x_{-i}(t)) \geq U_i\left(\exp[-h_g t]\frac{H}{x_i^{m_i^*}(0)}\prod_{j \neq i} x_j^{S_j}(0)\right)$$

Since $U_i\left(\frac{H}{x_i^{m_i^*}(0)}\prod_{j \neq i} x_j^{S_j}(0)\right) < \mu$, this implies by assumption d.3 that

$$(g_i(m_i^*; x_{-i}(t)) - g_i(m_i; x_{-i}(t))) > -\left(-\ln\left(U_i\frac{H}{x_i^{m_i^*}(0)}\prod_{j \neq i} x_j^{S_j}(0)\right) + h_g t\right)_{+}$$

So by integration we have that

$$\frac{x_i^{m_i^*}(t')}{x_i^{m_i}(t')} < L\frac{x_i^{m_i^*}(0)}{x_i^{m_i}(0)}$$

which is a contradiction. Since this exhausts all cases the result follows. 2

Proof of Proposition 5.

(i) We know, from Proposition 2, that $\dot{x}_3 > 0$ in any interior point. This implies that if there is a time t such that $x_3(t) > \frac{1}{2}$, then $x_3(t') > \frac{1}{2}$ for all $t' \geq t$. From equations (3-4) we have that, whenever $x_3(t) > \frac{1}{2}$, $\phi_i(x) > 0$ for players 1 and 2. This implies that if there is a time t such that $x_3(t) > \frac{1}{2}$, then $\dot{x}_i > 0$ for all $t' \geq t$ for $i = 1; 2$ and therefore x(t) converges. Since convergence must be to a Nash equilibrium and x_1 and x_2 have been increasing, x converges to NE^1. To show the stability of NE^1 it suffices to show that there is a neighborhood of NE^1 such that, for all x(0) in this neighborhood, there is a time t such that $x_3(t) > \frac{1}{2}$. Let $x_i(0) = 1 - \varepsilon_i$ for $i = 1; 2$ and $x_3(0) = \frac{1}{2} - \delta$, with $\varepsilon_i > 0$, $\delta > 0$. From (3-4) we also have that $-1 < \phi_i(x) < 1$ for $i = 1; 2$, thus

$$\exp[-t](1 - \varepsilon_i) < x_i(t) < \exp[t](1 - \varepsilon_i): \tag{6}$$

Since $\phi_3(x) \leq \frac{x_1(1-x_1)}{6}$ we have by equation (6)

$$\frac{\dot{x}_3(t)}{x_3(t)} > \frac{(1 - \varepsilon_1)(\exp[-t](1 - \exp[t](1 - \varepsilon_1))}{6}$$

thus

$$\frac{\dot{x}_3(t)}{x_3(t)} > \frac{(1 - \varepsilon_1)(\exp[-t] - (1 - \varepsilon_1)}{6} > \frac{(1 - \varepsilon_1)(-t + \varepsilon_1)}{6}:$$

This implies that

$$x_3(t) > \exp\left[\frac{(1 - \alpha_1)(-\frac{t^2}{2} + \alpha_1 t)}{6}\right](\frac{1}{2} - \pm):$$

Note that for $t = \alpha_1$

$$\exp\left[\frac{(1 - \alpha_1)(-\frac{t^2}{2} + \alpha_1 t)}{6}\right] \exp\left[\frac{(1 - \alpha_1)(\frac{\alpha_1^2}{2})}{6}\right] > 1$$

and therefore $x_3(t) > \frac{1}{2}$ for $\pm$ small enough, which is what we wanted to show.

(ii). Assume that $x_3(0) > \frac{3}{7}$. Since $\underline{x}_3(t) \geq 0$ for all t, $x_3(t)$ is an increasing function of t, therefore it must converge. Since the initial condition $x_3(0)$ is larger than $\frac{3}{7}$ it must converge to a number larger than $\frac{3}{7}$. We know that x(t) converges to a Nash equilibrium by Proposition 2. Since there is no equilibrium in NE^0 with $x_3 > \frac{3}{7}$, x(t) cannot converge to a point in NE^0. Since $x_3(0)$ can be arbitrarily close to $\frac{3}{7}$ and therefore to the set NE^0, this set must be unstable. 2

Proof of Proposition 6. The proof is constructed as follows. We first characterize the limit of the set of rest points $\hat{RE}(\eta)$, and then analyze the stability properties of each of its elements.

We start by observing that, given $\eta \in (0;1)$; any rest point must be completely mixed, and it also must be $x_3 > \eta$; as $\dot{x}_3(\cdot)$ is always positive in the interior of the state space Δ (because m_3^0 is a weakly dominated strategy).We also know, by the continuity of the vector field with respect to ε; that every limiting rest point of the dynamic, as ε goes to zero, must lie in the set of restpoints of the unperturbed dynamic $RE(\eta)$.

We analyze first the limit set of rest points under CASE 0. In this case, both players 1 and 2 play their strategy m_i^0 with probability 1, that is $x_i^0 = 0$; for $i = 1;2$. Setting $\underline{x}_1 = 0$ yields the following equation:

$$\frac{x_1}{\varepsilon} = \frac{12\left(\frac{1}{2} - x_1\right)}{(1 - x_1)(3 + x_1 - x_3(7 - x_2))} \tag{7}$$

and an analogous expression can be obtained for $\frac{x_2}{\varepsilon}$: Denote by x_3^0 a limiting value in a rest point, if a limit exists, for x_3. When the limiting values for x_1 and x_2 are zero we have:

$$\lim_{\substack{x_i \to 0 \\ \varepsilon \to 0}} \frac{x_i}{\varepsilon} = \frac{1}{2(3 - 7x_3^0)} \tag{8}$$

Notice that in this case if a rest point exists it must be $x_3^0 < \frac{3}{7}$; since $\frac{x_i}{\varepsilon} > 0$: We then set $\frac{x_3}{\varepsilon} = 0$;substitute $\frac{x_i}{\varepsilon}$ with the expression in (8), solve for x_3; and substitute $x_i; i = 1;2$ and ε by their limiting value of zero. The solutions for x_3^0 take the following form:

$$x_3^0 = \frac{1 + 7\eta + \sqrt{1 - \eta(46 - 49\eta)}}{10} \text{ and } x_3^0 = \frac{1 + 7\eta - \sqrt{1 - \eta(46 - 49\eta)}}{10}$$

Remember that x_3^0 must be a real, positive number, with $\bar{} < x_3^0 < \frac{3}{7}$. For the expression under the square root at the numerator to be nonnegative, it must be that $\bar{} \; 2 \; 0; \frac{23 \, i \, 4\sqrt{30}}{49} \, ¼ :0222673$; which determines the feasible range for both roots. Within this interval of values for $\bar{}$, x_3^0 (x_3^0) is a strictly decreasing (increasing) function of $\bar{}$; which has a minimum and a maximum, whose values are $\frac{15 \, i \, 2\sqrt{30}}{35}$ (0) and $\frac{2}{10} \quad \frac{15 \, i \, 2\sqrt{30}}{35}$ respectively. As $\bar{} \; ! \; \frac{23 \, i \, 4\sqrt{30}}{49}$; both solutions converge to $\frac{15 \, i \, 2\sqrt{30}}{35}$:

We now deal with the subset of limiting rest points under CASE 1, i.e. with limiting values for $x_i = 1$ for $i = 1; 2$. The equations corresponding to (7-8) are now the following:

$$\frac{(1 \, ¡ \, x_1)}{¸} = \frac{\frac{1}{2} \, ¡ \, x_1}{x_1 \; \frac{1}{3} + (1 \, ¡ \, x_3)(\frac{1}{12}(1 \, ¡ \, x_2) \, ¡ \, \frac{2}{3})} \tag{9}$$

$$\lim_{\substack{x_i \, ! \, 1 \\ ¸ \, ! \, 0}} \frac{(1 \, ¡ \, x_1)}{¸} = \frac{1}{2 \; \frac{1}{3} \, ¡ \, \frac{2}{3}(1 \, ¡ \, x_3)} \tag{10}$$

Denote by x_3^1 a limiting value in a rest point for x_3 in this latter case: By analogy with CASE 0, we know from (10) that, if a rest point exists, it must be $x_3^1 > \frac{1}{2}$: There is a unique feasible solution for $x_3^1; 8\bar{} \; 2 \; (0; 1)$ with the following form:

$$x_3^1 = \frac{3 + 4\bar{} + \sqrt{9 \, ¡ \, 16\bar{}(1 \, ¡ \, \bar{})}}{10}$$

Following the same procedure for the remaining rest points of the unperturbed dynamics (i.e. the pure strategy profiles which belong to RE ($¡$) and do not satisfy either CASE 0 or CASE 1) does not add any element to the limiting set of rest points of the perturbed dynamics. This should not be surprising, as any other rest point of the unperturbed replicator dynamics is unstable with respect to the interior. Since this exhausts all cases, the result follows.

We now move to establish the stability properties of each limiting restpoint separately: The Jacobian matrix for the dynamic system is as follows:

$$J(x; ¸) = \begin{matrix} (1 \, ¡ \, 2x_1)¢_1 \, ¡ \, ¸ & \frac{¡(1 \, ¡ \, x_1)x1(1 \, ¡ \, x_3)}{12} & \frac{(1 \, ¡ \, x1)x1(7+x_2)}{12} \\ \frac{¡(1 \, ¡ \, x_2)x_2(1 \, ¡ \, x_3)}{12} & (1 \, ¡ \, 2x_2)¢_2 \, ¡ \, ¸ & \frac{(1 \, ¡ \, x_2)x_2(7+x_1)}{12} \\ \frac{(1 \, ¡ \, 2x_2)(1 \, ¡ \, x_3)x_3}{6} & \frac{(1 \, ¡ \, 2x_1)(1 \, ¡ \, x_3)x_3}{6} & (1 \, ¡ \, 2x_3)¢_3 \, ¡ \, ¸ \end{matrix}$$

We analyze CASE 0 first. We know that, in this case, we have two restpoints, which we call $x^0 \; ´ \; (0; 0; x_3^0)$ and $x^0 \; ´ \; (0; 0; x_3^0)$: We evaluate the Jacobian when $x_1; x_2$ and $¸$ are equal to their limiting value (i.e. 0). The corresponding eigenvalues are: $0; \frac{¡3+7x_3^0}{12}; \frac{¡3+7x_3^0}{12}$: There are then two (identical) negative eigenvalues (since any limiting $x_3^0 < \frac{3}{7}$ for CASE 0), while the third eigenvalue is equal to zero. To determine the stability properties of the perturbed system, the sign of the eigenvalue whose limit is zero becomes crucial given that the continuity of $J(:)$ ensures that the other two will be negative, for any $¸$

su±ciently small. We now linearize the rest points (as a function of ¸) around NE^0. We set $x^*(¸;±) ´ (±_1¸;±_2¸;x_3^0 + ±_3¸)$; where $± ´ (±_1;±_2;±_3)$ denotes the vector collecting the coe±cients of the linearised system. We then evaluate the following expression:

$$Á^0(x_3^0;±) ´ \lim_{¸ \to 0} \frac{@ \det \bar{J}(x;¸)_{x^*(¸;±)}}{@¸}$$

We do so because $\det(J(x;¸))$; which is equal to zero $8x \in NE^0$; will preserve the sign of the third eigenvalue, given that the sign of the other two will stay constant (and negative) when x is su±ciently close to NE^0 and ¸ is su±ciently small. For CASE 0 we get the following result:

$$Á^0(x_3^0;±) = \frac{¡54 + x_3^0(252 + 294x_3^0) + (±_1 + ±_2)(9 ¡ 39x_3^0 + 63(x_3^0)^2 ¡ 49(x_3^0)^3)}{864} \tag{11}$$

We ¯rst notice that (11) does not depend on $±_3$. To evaluate $\text{sign}(Á^0(x_3^0;±))$we only need to get estimates of $±_1$ and $±_2$, the linear coe±cients which measure the responsiveness of the equilibrium values of $x_i; i = 1;2$ to small changes in ¸. We do so setting $\lim_{¸ \to 0} \frac{@}{@¸} \bar{D}(x;¸)_{x^*(¸;±)} = 0$ and solving for $f±_1;±_2;x_3^0g$: There are two alternative set of solutions, each of them corresponds to each of the restpoints. In particular:

$$±_1^0 = ±_2^0 = \frac{23 ¡ 49¯ ¡ 7\sqrt{1 ¡ ¯(46 ¡ 49¯)}}{8}$$

$$\hat{±}_1^0 = \hat{±}_2^0 = \frac{23 ¡ 49¯ + 7\sqrt{1 ¡ ¯(46 ¡ 49¯)}}{8}$$

We evaluate the numerator of (11) for both sets of solutions, and we get the following expressions:

$$Á(¯) = \frac{3(¡7 + 322¯ ¡ 343¯^2 + (49¯ ¡ 23)\sqrt{1 ¡ 46¯ + 49¯^2})}{10} \tag{12}$$

$$\hat{Á}(¯) = \frac{2863 ¡ 147476¯ + 882882¯^2 ¡ 1546244¯^3 + 823543¯^4 + k\sqrt{1 ¡ 46¯ + 49¯^2}}{1000} \tag{13}$$

with $k = (3887 ¡ 60123¯ + 165669¯^2 ¡ 117649¯^3)$:

Both $Á^0(¯)$ and $\hat{Á}^0(¯)$ are plotted in Figure 4. As the diagram shows, $Á^0(¯)$ is always negative in the domain $[0; \frac{23¡4\sqrt{30}}{49})$; whereas $\hat{Á}^0(¯)$ is not. As a result of that, x^0 is asymptotically stable whereas $\hat{x}^0$ is not.

Figure 4

Asymptotic stability of the dynamic with drift

We now move on to CASE 1. Here we have a unique rest point, which we call $x^1 = (1; 1; x_3^1)$: The eigenvalues of the unperturbed dynamics are as follows: $0; \frac{1-2x_3}{3}; \frac{1-2x_3}{3}$: As in CASE 0, there are two (identical) negative eigenvalues (given that $x_3 > \frac{1}{2}$); and the remaining eigenvalue equal to zero. By analogy with CASE 0, we now define $x(\varepsilon) = (1 - \pm_1\varepsilon; 1 - \pm_2\varepsilon; x_3^0 + \pm_3\varepsilon)$ and solve $\lim_{\varepsilon \to 0} \frac{\partial}{\partial \varepsilon} D(x; \varepsilon)_{x(\varepsilon;\pm)} = 0$ to get estimates of $\pm$: The unique feasible solution (corresponding to the unique limiting equilibrium), takes the following form:

$$\hat{\pm}_1^1 = \hat{\pm}_2^1 = \frac{3 - 2 - 4\bar{} + \sqrt{9 - 16\bar{} + 16\bar{}^2}}{2}$$

The function corresponding to (12-13) takes now the following form:

$$\hat{A}^1(\bar{}) = \frac{24\,(-\circledR + (2 - 4\bar{})\sqrt{\circledR})}{5}$$

with $\circledR = 9 - 16\bar{}$: The function $\hat{A}^1(\bar{})$ is also plotted in Figure 4. As the diagram shows, $\hat{A}^1(\bar{})$ stays negative $8\bar{} \in (0; 1)$: As a result of that, x^1 is asymptotically stable under any drift configuration. 2

$m_3 = m_3^0$

	m_2^0	m_2^1
m_1^0	$\frac{1}{4},\frac{1}{4},\frac{1}{2}$	$\frac{1}{3},0,\frac{1}{3}$
m_1^1	$0,\frac{1}{3},\frac{1}{3}$	$0,0,\frac{1}{3}$

$m_3 = m_3^1$

	m_2^0	m_2^1
m_1^0	$0,0,\frac{1}{2}$	$0,\frac{1}{3},\frac{1}{2}$
m_1^1	$\frac{1}{3},0,\frac{1}{2}$	$\frac{1}{3},\frac{1}{3},\frac{1}{3}$

Figure 1

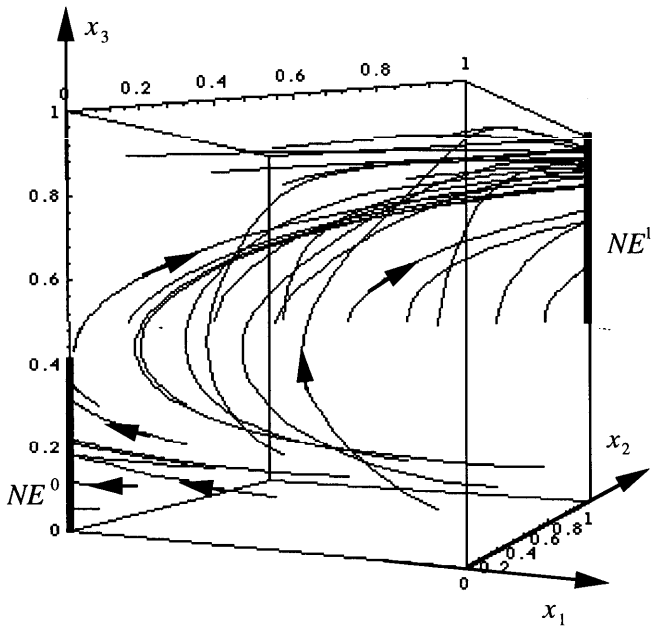

Figure 2

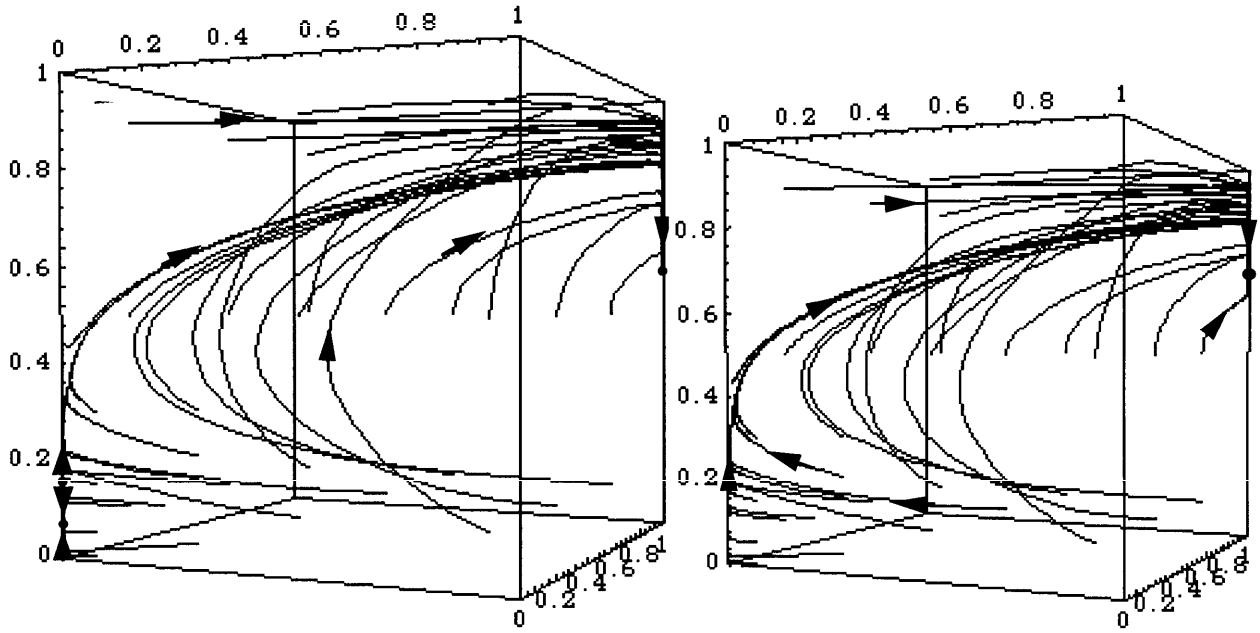

Figure 3

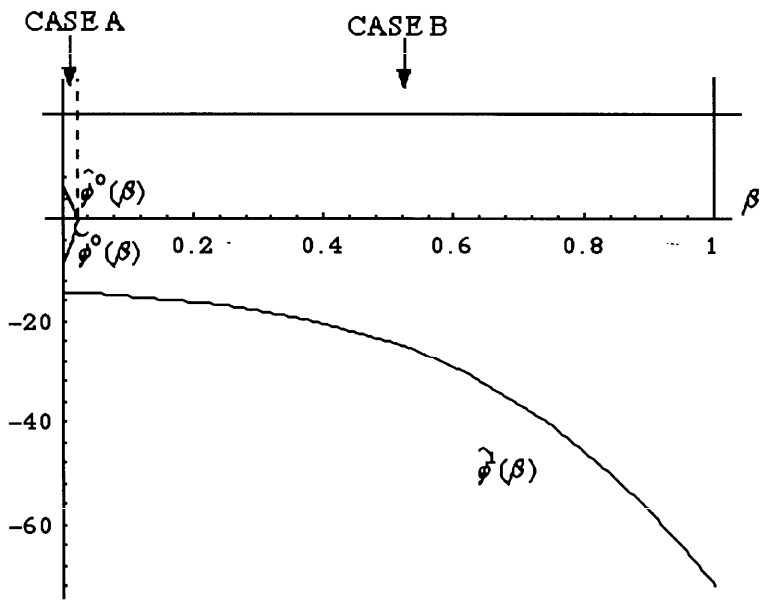

Figure 4