Aspirations, adaptive learning and cooperation in repeated games

Bendor, J.; Mookherjee, D.; Ray, D.

Publication date:
1994

entER
for
omic Research

Discussion paper

TILBURG UNIVERSITY

Center
for
Economic Research

No. 9442

ASPIRATIONS, ADAPTIVE LEARNING AND COOPERATION IN REPEATED GAMES

by Jonathan Bendor, Dilip Mookherjee,
and Debraj Ray

May 1994

ISSN 0924-7815

Aspirations, Adaptive Learning and Cooperation in Repeated Games[1]

Jonathan Bendor, Graduate School of Business, Stanford University
Dilip Mookherjee, Indian Statistical Institute, New Delhi
Debraj Ray, Department of Economics, Boston University

revised, March 1994

Abstract

We consider a model of a two player repeated game with "satisficing" players who choose actions based on their experience in past plays. A player learns adaptively by increasing probability weight on a recently chosen action if it realized a payoff above an aspiration level, and decreasing it otherwise. Aspirations in turn are required to be consistent with the long run average payoffs induced by this process. Our equilibrium notion incorporates stability under small "trembles" by players. It is shown that (i) non-Nash behavior may result in the long run, (ii) players can generally attain individually rational *pure action* Pareto efficient payoffs, (iii) multiple long run equilibria arise in some contexts, hence initial conditions may matter, (iv) inefficient mixed strategy payoffs cannot generally be achieved, and (v) in specific games such as coordination games and the Prisoners' Dilemma the set of long run equilibria can be narrowed down substantially.

1 Introduction

We consider a model of "satisficing" behaviour by two players engaged in a repeated game.[2] An action is deemed "satisfactory" if its current payoff exceeds some *aspiration* level held by the player. Each player is assumed to adapt his strategy from one iteration to the next by increasing the probability weight on a recently chosen

[1]An earlier version of this paper was written while the second author was visiting the Center for Economic Research, Tilburg University in October 1991. We would like to thank Kenneth Arrow, Eric van Damme, Curtis Eaton, Chris Harris, Matt Jackson, Charlie Kahn, Roger Myerson, Rafael Rob, Avner Shaked, Vernon Smith, Jeroen Swinkels, Fernando Vega Redondo and Jorgen Weibull for useful comments and discussions. The paper has also benefitted from presentations at Tilburg, Illinois, Northwestern, Delhi School of Economics, Studienzentrum Gerzensee, the 1993 Berkeley Decentralization Conference, and the Tenth World Congress of the International Economic Association, Moscow.

[2]For discussion of the notion of "satisficing" in relation to traditional economic models of rational choice, see Simon (1955, 1957, 1959) and March and Simon (1958). It also played an important role in the evolutionary theory of Winter (1971) and Nelson and Winter (1982). A recent axiomatic theory with similar features appears in the work of Gilboa and Schmeidler (1992).

action if it was satisfactory, and decreasing it otherwise. Morever, aspirations levels are themselves endogenous: they must be consistent with long run average payoffs. Finally, long run outcomes are required to be stable with respect to "small trembles" that correspond to occasional lapses of memory, or replacement of players by new ones.

This model extends theories of learning formulated by mathematical psychologists in the 1950s, especially Estes (1954) and Bush and Mosteller (1955). Such theories are supported by a substantial amount of experimental evidence.[3] Our model embeds a similar learning process in a two player repeated game, where the players revise their strategies following every interaction.

In a sense, the interaction pattern studied in this paper is a polar opposite to the random matching context considered by most of the literature on evolution and learning. In most evolutionary models, pairs of players are selected randomly from a "large" population to play the given game once, and are thereafter returned to the population. In learning models, each player is assumed to interact with many varying opponents with a fixed strategy, modifying his strategy thereafter based on the cumulative experience. In both classes of models, the "fitness" of a given strategy at any stage of the game depends on its average payoff, achieved against the rest of the population.[4] In such a context, a strategy revision by any single player evokes no response from the other players, as it does not appreciably affect their average payoffs. In contrast, the strategy revisions of a given player in our model generate substantial feedback effects by affecting the other player's payoffs, thereby inducing the latter to also revise his strategy subsequently.

The combination of aspiration-based learning with repeated interaction between a small number of players in our model gives rise to several distinctive features:

1. The process may converge to a pure strategy limit which is *not* a Nash equilibrium of the one-shot game. This is due to the inter-player feedback effects that do not allow players to sustain payoff gains from unilateral deviations for any appreciable length of time.[5]

2. Players frequently learn to "cooperate", as the feedback effects turn out to resemble the outcome of "trigger strategies". Indeed, any individually rational, Pareto efficient *pure strategy* outcome can constitute a stable long-run outcome.

3. Since individual learning is dependent on players' aspiration levels, which are determined endogenously, there may be multiple long run equilibrium outcomes associated with different aspiration levels. Hence initial conditions can matter, even in the presence of small amounts of random "trembles" or "mutations".

4. Nevertheless, the set of long run equilibrium outcomes is relatively small, especially in contrast to Folk-Theorem-type results. For instance, mixed strategy

[3] See, for instance, Bush and Mosteller (1955); Suppes and Atkinson (1960); Selten and Stoecker (1985); Mookherjee and Sopher (1992); and Roth and Erev (1993).

[4] For a useful discussion of these models, see Binmore and Samuelson (1993).

[5] For further discussion, see Section 2.

outcomes that are Pareto dominated by some pure strategy outcome generally do not constitute long run equilibria. In coordination games, or the Prisoners' Dilemma, the set of long run equilibria can be narrowed down to a small number; for some parameter values the long run equilibrium is unique.

The following Section provides a more detailed overview of our model, and discusses the relation to existing literature. Section 3 presents the formal model, and Section 4 establishes a benchmark result concerning long run outcomes induced by given aspiration levels when players do not tremble. Section 5 presents some general results for equilibrium outcomes that are robust to trembles, while Section 6 discusses predictions generated for a coordination game and the Prisoners' Dilemma. Section 7 concludes.

2 Overview of Our Model

We first provide an informal discussion of the learning rule used by players in our model. The *state* of any player at any stage t is represented by a probability vector over his set of pure actions. One may interpret these probabilities as representing a player's relative inclination to select different actions: in the spirit of stochastic choice theory, this presumes the presence of other unmodelled determinants of a player's *actual* choices. A player's state is updated in the following manner: if the payoff realized from the action chosen at t exceeds an aspiration level, the weight on that action is increased at the following stage, with compensating adjustment in the weights on other available actions. Conversely, if the achieved payoff at t falls short of the aspiration level, then other actions will be tried with positive probability at the following iteration: this is, of course, consistent with a reduction in the weight on the previously chosen action. As defined, we allow for the possibility that a player's state will remain unchanged if the achieved payoff exactly equals the aspiration level.

The aspiration level, and therefore the learning rule of each player is assumed to be fixed throughout the duration of the game. Psychological evidence suggests, however, that aspiration levels themselves adapt to experience.[6] Nevertheless, aspirations adapt "more slowly" to experience than do actual strategies.[7] There are various ways of modelling the exact process by which aspirations adapt to experience. For instance, one might set the aspiration level at any stage equal to the time average of payoffs achieved in past plays, or more generally, to some convex combination of such payoffs. Alternatively, each player (or overlapping generations of players) could be involved in a succession of repeated games, where each constitutent repeated game is played with a given aspiration level that is modified from one repeated game (or generation) to the next based on the intervening payoffs.

In this paper we do not model the dynamics of aspiration levels, *i.e.*, the evolution of the learning rules themselves. Instead, we explore long run equilibrium notions

[6] See Simon (1959).

[7] Throughout this paper, we use the terms "strategy" and "action" interchangeably: there is no notion of strategy here in the sense of a history-dependent policy.

where aspiration levels equal long run average payoffs associated with the learning process induced by these aspirations. In other words, we focus on how players learn to play a game with fixed aspirations, which in turn have a "self-fulfilling" property. This may be viewed as a useful first step in a more complete theory of learning, which would allow for changing aspirations in the course of the game, and might also specify how players revise their learning rules. One may view our equilibrium notion as characterizing the stationary points of such a model.

Despite the endogeneity of aspiration levels, there is scope for considerable multiplicity of equilibria. In particular, any pure action vector can be rationalized as an equilibrium associated with aspirations equal to the corresponding payoffs: if initial aspirations exactly equal these payoffs, neither player might be motivated to move away from these strategies. It therefore makes sense to require long run equilibria to be stable with respect to small random perturbations, or *trembles*, of the state of either player. Two different stability notions are studied. An *equilibrium with nearly consistent aspirations (ENCA)* is a distribution D^* over different (mixed) actions of players which is a limit distribution of the stochastic learning process induced by certain aspiration levels, satisfying two conditions: (a) the aspiration of each player equals his expected payoffs under D^*; and (b) D^* is the limit of a sequence of ergodic distributions D^v generated by a sequence of learning processes with the *same* aspirations, but vanishingly small trembles. The qualification "nearly" comes from the fact that aspirations need not equal limit expected payoffs under the trembled processes, equality only being demanded as trembles vanish. A stronger notion of stability is embodied in the notion of an *equilibrium with consistent aspirations (ECA)*, addresses exactly this point. Under an ECA, the limit aspirations must also be a limit of aspirations along the trembled processes, and these trembled aspirations must equal expected payoffs too.[8]

The "satisficing" model of strategy learning can be contrasted to "myopic best-response" models studied in recent literature, such as "fictitious play" or related models studied by Robinson (1951), Shapley (1964), Fudenberg and Kreps (1988), Canning (1989, 1991), Jordan (1991), Milgrom and Roberts (1991), Krishna (1992), Kandori, Mailath and Rob (1993) and Young (1993) among others. The difference between the two approaches has been discussed by Selten (1991), under the heading of "stimulus learning" and "belief learning" respectively. In the latter, each player observes the past moves of other players, forms beliefs about their strategies in the next iteration of the game, and then selects a best response. This approach requires players to (i) know their own payoff function, (ii) form beliefs about opponents' choices at the next iteration based on choices observed in the past, and (iii) calculate an optimal response. It therefore presumes that players are able to articulate a model of their environment, to collect and process information about the choices made

[8] Yet stronger would be a requirement of stability against *all* possible sequences of experimentation, rather than one particular sequence, analogous to the distinction between the notion of strategic stability of Kohlberg-Mertens (1986) and of trembling-hand-perfect equilibrium of Selten (1975). The characterization theorems of Section 5 would continue to hold for such a notion, but the existence of such equilibria cannot be guaranteed generally.

by others; compute expected payoffs corresponding to different actions available, and select a maximizing response.[9] The "stimulus learning" approach on the other hand applies to players ignorant of payoff functions and of opponents' past choices. It does not require them to solve maximization problems. The two approaches therefore presume different levels of bounded rationality, characterized by different limits on information gathering or cognitive abilities.[10] Ultimately, of course, the relative appropriateness of different models (e.g., "belief learning" versus "stimulus learning") is an empirical matter, on which experiments ought to throw some useful light.[11]

Even within the class of "satisficing" hypotheses, different modeling alternatives arise. For instance, obtaining a payoff above aspiration may leave the current strategy unaltered, while a payoff below aspiration may trigger a different strategy. Or a player could keep track of past time average payoffs associated with different strategies chosen previously, and allocate weight based on these average scores, as in the single agent model of Arthur (1993). This would imply that players would learn "less" with greater experience. Each of these variations would induce a different dynamic. The particular formulation adopted in this paper is based on the Bush-Mosteller hypothesis which has received considerable support in the experimental psychology literature. It does not presume, however, that this particular formulation is intrinsically better than the others on either *a priori* or experimental grounds. In our view, exploring the sensitivity of our results to variations in the specification of satisficing behavior is an important task for future research.

The importance of the "matching" framework also warrants emphasis. In conjunction with the satisficing hypothesis, it is essential to many of our results. Repeated interaction between a few players generates significant feedbacks in their strategy revisions. This is illustrated by our observation that stable long run outcomes need not be Nash equilibria of the one-shot game. Indeed, feedback effects operate in a manner that resembles "punishments" imposed on unilateral deviations in repeated games.

Consider, for example, the Prisoners Dilemma. It turns out that mutual co-

[9]However, some of these "best response" dynamic models are consistent with "less rational" behavior, such as imitation of successful strategies in the context of random matching of pairs from a large population, as discussed by Kandori, Mailath and Rob (1993). Useful results can also be obtained for specific games such as supermodular games on the basis of weak assumptions concerning belief formation processes, as in Milgrom and Roberts (1991).

[10]See Simon (1955, 1959) and Selten (1978, 1991) for interesting discussions of different degrees of "bounded rationality".

[11]Selten and Stoecker (1986) describe an experiment on the finitely repeated Prisoners' Dilemma, in which the nature of play is predicted rather well by a parsimonious model of the Bush-Mosteller variety. In constant-sum two player repeated game experiments, Mookherjee and Sopher (1992, 1993) find that stimulus learning satisfactorily fits the pattern of play when players are not informed of the payoff matrix or opponents' choices. When players are fully informed, however, then fictitious play or similar hypotheses describe the pattern of play better if the game is sufficiently "complex" (measured by the number of pure strategies available to each player), but not otherwise. Roth and Erev (1993) discuss the usefulness of Bush-Mosteller-type models in explaining the results of experiments with different two stage sequential games.

operation is always an ECA (and therefore ENCA) outcome, despite the fact that mutual defection is a dominant strategy equilibrium in the one shot game. What prevents the cooperators from "learning" the payoff advantage to defection? Start with aspirations near the cooperative payoff. Suppose that player 1 experiments with defection at stage t. Since player 2 continues to cooperate, player 1 obtains a payoff higher than his aspiration, thereby making him even more inclined to deviate at $t+1$. Player 2 however ended up with a payoff *below* at t: this *also* makes 2 more inclined to defect at $t+1$, though for entirely different reasons. Suppose, then, that both defect at $t+1$. *Then both players receive below aspiration payoffs at* $t+1$*, thus tending to motivate both to return to cooperation at* $t+2$. Thus, once the players arrive at a state where both defect with substantial probability, both will indeed defect simultaneously, beginning the process of a simultaneous return to cooperation. Hence, the mutual cooperation outcome is stable with respect to periodic random switches to defection by either player. Morever, the observed pattern of play will resemble that of sophisticated players using "trigger strategies".

In the random matching framework, such an outcome could not survive in the long run: a deviation by a single player will not have an appreciable impact on the "fitness" of any other strategy in the population, and therefore not evoke any feedback effect. The initial benefits of the deviation are then not "undone" by the reactions of other players, and the player can sustain the benefits of the unilateral deviation. Consequently, the fitness of the deviating strategy will be enhanced, at the eventual expense of other strategies in the population.

Finally, we discuss the relation of our approach to the evolutionary theory of Winter (1971) and Nelson and Winter (1982). In a model of selection among different production activities (as well as their scale of operation), they explore the implications of production choice rules that adapt choices to current profitability: profitable activities are expanded, losing activities are contracted and activities that just break even retain their current size. (In our framework, this corresponds to a Bush-Mosteller adaptive rule combined with an aspiration level of zero profits.) In the presence of "innovating remnants", which introduces a certain amount of random "trembles", they show that the process converges to a competitive equilibrium (where firms earn their aspiration profit levels of zero). Our approach is similar in that different players adapt without explicitly coordinating with one another and without devoting any cognitive effort to predicting the choices of others and choosing appropriate responses to such predictions.

3 The Model with Fixed Aspirations

There are two players A and B, with finite action sets $\mathcal{A}$ and $\mathcal{B}$. A typical pure strategy will be denoted by $a \in \mathcal{A}$, $b \in \mathcal{B}$. Let $\mathcal{C} \equiv \mathcal{A} \times \mathcal{B}$; a *pure strategy profile* is then a pair $c = (a, b) \in \mathcal{C}$. Player A has a *payoff function* $f : \mathcal{C} \to \mathbb{R}$ and B has a payoff function $g : \mathcal{C} \to \mathbb{R}$.

A *mixed strategy* for A (resp. B) is an element α (resp. β) of $\Delta(\mathcal{A})$ (resp. $\Delta(\mathcal{B})$), where $\Delta(X)$ denotes the set of all probability measures on a set X. Clearly,

$\Delta(\mathcal{A})$ and $\Delta(\mathcal{B})$ are finite dimensional unit simplices. Let $\gamma \equiv (\alpha, \beta)$ denote a *mixed strategy profile*. Then γ is a product probability on $\mathcal{C}$. Let S be the space of all such mixed strategy profiles: $S = \Delta(\mathcal{A}) \times \Delta(\mathcal{B})$. Finally, the *aspiration levels* of A and B are denoted by numbers F and G respectively.

3.1 Adaptive Learning Rules

For A, let α be an ongoing strategy, a an action chosen, and f a payoff received from choosing a. An *adaptive learning rule* maps this information (and the aspiration level for A) into a new mixed strategy $L^A(\alpha, a, f, F)$ in $\mathcal{A}$. A similar definition holds for the learning rule L^B used by player B. We thus allow for the possibility that a player may wish to "hedge" between different actions, or may not be sufficiently inclined on the basis of past experience to select one action deterministically. The experience from playing any particular round causes a player to adjust the probability weights on different actions, but not necessarily switch actions altogether. Specific restrictions on such adjustments will be introduced in Section 4.

Where there is no confusion, we will denote by α', β', α_{t+1}, β_{t+1}, etc. the mixed strategies "in the next period" induced by the learning rule.

3.2 Learning with Trembles

We will permit players to occasionally "start over" again, in a sense to be made precise below. Several well known interpretations of this postulate are possible: players occasionally forget the past, they experiment, they err, they die and are replaced by others with the same payoffs and aspirations but no knowledge of the past, and so on. We will use the term "trembles" to refer to these occasional changes.

Specifically, we assume that at each date, with some (independent) probability, each player restarts the learning process by choosing, according to some exogenously given measure, a mixed strategy from the set of all mixed strategies. This forms the starting point of a fresh learning process.

To precisely define the trembling process, and for ease of later exposition, we describe (for individual A) the order in which events unfold. At any date, A is equipped with a particular mixed strategy α. This strategy has presumably been built up via adaptive learning (and/or trembles) in the past. At the current date, this strategy α is used to choose a particular *action*, and a payoff f is received, based also on the action simultaneously chosen by the other player.

At this stage, the learning rule L^A moves A to a new mixed strategy. However, A might not use the learning rule. With a small exogenous probability ϵ, player A might "tremble", or "start all over again". In this event, A selects a new mixed strategy according to some exogenously given measure ν^A over the space $\Delta(\mathcal{A})$ of all his mixed strategies. So that any mixed strategy may be the outcome of the tremble, we assume that ν^A has full support on $\Delta(\mathcal{A})$. In any case, the player will possess a new mixed strategy α' at the start of the next period, and the same story repeats itself there onwards. We shall essentially be interested in the case where the tremble probability ϵ becomes vanishingly small, as will become clear below.

3.3 The Adaptive Learning Process

Our model describes a Markov process on the state space S. Recall that S is the set of all mixed strategy profiles γ in $\mathcal{A} \times \mathcal{B}$. Give S the Borel σ-algebra. Fix aspirations (F, G), learning rules (L^A, L^B), and tremble probability ϵ (assumed common for both players, without any essential loss of generality). For each mixed strategy profile γ in S, a transition probability $P(\gamma, .; F, G, \epsilon)$ can be defined as follows. Given $\gamma = (\alpha, \beta)$, the pure action outcome $c = (a, b)$ is realized with probability

$$p(\gamma, c) \equiv \alpha(a)\beta(b) \tag{1}$$

This outcome causes individual A to pass to a new mixed strategy over his actions. The passage occurs *either* via adaptive learning (*i.e.* using the rule $L^A(\alpha, a, f(c), F)$, absent any tremble), *or* via a tremble (whence a new mixed strategy is selected according to the "tremble measure" ν^A on his mixed strategies). A similar description applies to individual B.

So if $\mathcal{L}^A(.; \epsilon)$ and $\mathcal{L}^B(.; \epsilon)$ denote the resulting measure on A and $B's$ mixed strategies in the following period, we have (where $\delta\alpha$ and $\delta\beta$ denote the degenerate measures concentrated at mixed strategies α and β respectively):

$$\begin{aligned} \mathcal{L}^A(\alpha, a, f(c), F, \epsilon) &= (1-\epsilon)\delta L^A(\alpha, a, f(c), F) + \epsilon\nu^A \\ \mathcal{L}^B(\beta, b, g(c), G, \epsilon) &= (1-\epsilon)\delta L^B(\beta, b, g(c), G) + \epsilon\nu^B \end{aligned} \tag{2}$$

and outcome c therefore induces the product measure

$$Q(\gamma, c; F, G, \epsilon) \equiv \mathcal{L}^A(\alpha, a, f(c), F, \epsilon) \times \mathcal{L}^B(\beta, b, g(c), G, \epsilon) \tag{3}$$

on the state space S. We may combine (1) and (3) to yield a transition probability as

$$P(\gamma, S'; F, G, \epsilon) \equiv \sum_{c \in \mathcal{C}} p(\gamma, c) Q(\gamma, c; F, G, \epsilon)(S') \tag{4}$$

for any Borel subset S' of the state space S. This transition probability is parametrized by the aspiration levels F, G and the tremble probability ϵ.

Start with any initial measure μ_0 on S. Then given a pair of aspiration levels and a tremble probability, the transition probability described above determines a sequence $\{\mu_t\}$ of probability measures on S. For convenience of exposition, we shall hereafter use the terms *measure* and *distribution* interchangeably.

3.4 A Preliminary Result

We use standard arguments to establish that when players tremble with strictly positive probability, then for any initial distribution μ_0 over the set of mixed strategy profiles, the resulting sequence of distributions converges in a strong sense (and at a geometric rate) to a unique limit distribution. In other words, the long run outcome of the process is well defined, given any pair of aspiration levels and a positive tremble probability.

THEOREM **3.1** *For each pair of aspiration levels F, G and every tremble probability $\epsilon > 0$, there exists a unique invariant measure $\mu(F, G, \epsilon)$ such that for each initial μ_0 on S, μ_t converges to $\mu(F, G, \epsilon)$ as $t \to \infty$. Convergence occurs (geometrically) in the total variation norm (on S) and therefore also in the topology of weak convergence on S.*

Proof. For ease of notation, drop the explicit dependence on (F, G, ϵ) and let $P(.,.)$ denote the transition probability, with $P^N(.,.)$ its N-step extension. Using Theorem 11.12 of Stokey and Lucas [1989], it suffices to show that the following condition is met:

Condition M: *There exist $\epsilon' > 0$ and an integer $N \geq 1$ such that for every Borel subset S' of S, either $P^N(\gamma, S') \geq \epsilon'$ for all $\gamma \in S$, or $P^N(\gamma, S \setminus S') \geq \epsilon'$ for all $\gamma \in S$.*

Condition M is readily verified in the present context. Let $\epsilon' = \epsilon^2/2$, where ϵ is the tremble probability. Let N equal 1. Observe from (2) and (3) that for any $\gamma \in S$, any outcome $c \in C$ and any Borel subset S' of S, $Q(\gamma, c)(S') \geq \epsilon^2 \nu(S')$, where ν is simply the product measure $\nu^A \times \nu^B$. Consequently, by (4),

$$P(\gamma, S') \geq \epsilon^2 \nu(S')$$

for all γ and S'. Note that $\max\{\nu(S'), \nu(S \setminus S')\} \geq 1/2$. It follows, then, that for each Borel set S', either $P(\gamma, S') \geq \epsilon^2/2$ for all γ, or $P(\gamma, S \setminus S') \geq \epsilon^2/2$ for all γ. This verifies Condition M, and completes the proof. ■

4 Consistent Aspirations

So far we have described a situation where aspirations are not permitted to vary. It is possible to drop this restriction in different ways. A minimal requirement for a model of endogenous aspirations is that in the long run, aspirations should not be out of line with the average payoffs accumulated from experience. Otherwise one would expect players to modify their aspiration levels. Theorem 3.1 ensures that with positive trembles, the long run distribution over mixed strategy profiles is uniquely defined, and well-known ergodicity arguments imply that the time averages of payoffs are equal to the expected payoffs in the long run distribution. Hence it is natural to impose the requirement that aspiration levels equal the expected payoffs under the limit distributions (of the processes induced by those aspiration levels).

An *equilibrium with consistent aspirations* (ECA) is a probability distribution μ^* over S and associated aspirations F^* and G^* with the property that there exist sequences of strictly positive tremble probabilities $\epsilon_k \to 0$, aspirations (F_k, G_k) and probability distributions μ_k over S such that

(1) μ_k is the limit distribution under (F_k, G_k, ϵ_k) and (F_k, G_k) *equals* the expected payoffs (of A and B respectively) under the distribution μ_k.

(2) μ_k converges (weakly) to μ^* and (F_k, G_k) converges to (F^*, G^*) as $\epsilon_k \to 0$.

Why do we complicate an otherwise natural definition by invoking a sequence of (vanishing) trembles? The reason for this is straightforward: without trembles, *any* pure strategy pair might trivially be achieved as the long-run outcome of a game where players' average payoffs equal their aspiration levels. To elaborate, if players start with aspirations that equal the payoff from a pure action pair, and from a state where they play this pure action pair with probability one, then they might have no reason to alter their state thereafter, and will earn payoffs that equal their aspirations in every stage.[12] They may thus be "stuck" with actions that generate vastly inferior payoffs compared to other feasible strategies. Such outcomes would not be stable under small trembles, whence superior strategies would eventually be discovered and adopted.

Our equilibrium notion thus requires a pattern of long run behavior induced by aspirations which satisfy not only the consistency requirement that these aspirations equal the resulting average payoffs, but also impose stability with respect to small tremble probabilities.

As we have recognized in the preceding section, it is possible to demand more from a theory of endogenous aspirations. In particular, one might require that aspirations be continually modified *as the game proceeds*. Our model does not incorporate these more sophisticated notions. Observe, however, that even in this case, one would "ultimately" ask for equality of aspirations with long-run rewards. If these converge, our equilibrium notion may be viewed as describing the steady states of such extended models.

Under the following continuity assumption on learning rules, an ECA always exists. (From now on, we adopt the convention of stating all assumptions for player A, but take it that their obvious analogues hold for B as well.)

(A1: continuity in aspirations and initial state) *For each action a and payoff f, the mixed strategy induced by the adaptive learning rule $L^A(\alpha, a, f, F)$ is continuous in (α, F).*

THEOREM **4.1** *Under (A1), an equilibrium with consistent aspirations exists.*

Proof of Theorem 4.1. We shall use the following lemma, which we state separately for use in proving other results.

LEMMA **4.1** *Under (A1), the limit measure is continuous (in the weak topology) in (F, G, ϵ) for all $(F, G) \in \mathbb{R}^2$ and all $\epsilon > 0$. In addition, suppose that $(F_n, G_n, \epsilon_n) \to (F, G, 0)$. Then every limit point of $\mu(F_n, G_n, \epsilon_n)$ is an invariant measure for the adaptive learning process with $\epsilon = 0$.*

Proof. First we note, using (1)–(4) and (A1), that the transition probability $P(\gamma, ., F, G, \epsilon)$ defined in (4) in continuous in (γ, F, G, ϵ) for all $\gamma \in S$, $(F, G) \in \mathbb{R}^2$, and $\epsilon \in [0, 1]$. The details of this verification are standard and therefore omitted.

[12] This observation is, indeed, consistent with the assumptions that we shall place on learning rules; see below.

The first part of the lemma now follows from Theorem 12.13 in Stokey and Lucas [1989]. For the proof of the second part, all conditions of that theorem barring condition (c) continue to be met (at $\epsilon = 0$, a unique limit measure is *not* guaranteed). Nevertheless, if the last four sentences of their proof are omitted, the remainder suffice exactly for a proof of the second part of our lemma. ■

To continue the proof of the theorem, consider a compact rectangle K in $\mathbb{R}^2$ such that $(f(c), g(c)) \in K$ for all $c \in C$. Note that for each $\gamma \in S$, the functions $\hat{f}(\gamma) \equiv \sum_{c \in C} f(c)\gamma(c)$ and $\hat{g}(\gamma) \equiv \sum_{c \in C} g(c)\gamma(c)$ are continuous in γ (in the Euclidean topology), with image in the bounded set K.

For each μ on S, define $\hat{F}(\mu) \equiv \int \hat{f}(\gamma) d\mu(\gamma)$ and $\hat{G}(\mu) \equiv \int \hat{g}(\gamma) d\mu(\gamma)$. These are the expected payoffs to A and B respectively under the probability μ on S. By the observation in the previous paragraph, $\hat{F}(\mu)$ and $\hat{G}(\mu)$ are continuous on $\Delta(S)$ (in the topology of weak convergence), and assume values in K.

So, by the continuity properties established above, for given $\epsilon > 0$, the map $(F, G) \mapsto \mu(F, G, \epsilon) \mapsto \left(\hat{F}(\mu(F, G, \epsilon)), \hat{G}(\mu(F, G, \epsilon))\right)$ is continuous from K to K (in the Euclidean topology). By Brouwer's fixed point theorem, there exists a probability measure $\mu^*(\epsilon)$ and aspirations $(F^*(\epsilon), G^*(\epsilon))$ such that

$$F^*(\epsilon) = \hat{F}(\mu^*(\epsilon)), \quad g^*(\epsilon) = \hat{G}(\mu^*(\epsilon)) \tag{5}$$

and

$$\mu^*(\epsilon) = \mu(F^*(\epsilon), G^*(\epsilon), \epsilon). \tag{6}$$

To complete the proof, take any sequence $\epsilon_k \to 0$ and the corresponding sequence $(F^*(\epsilon_k), G^*(\epsilon_k), \epsilon_k)$ given by (5) and (6). Since S is compact, $\Delta(S)$ is compact in the weak topology, this sequence has a convergent subsequence, whose limit must be an ECA. ■

We close this section by considering a weaker notion of consistency in aspirations.

The notion of an ECA is demanding in that it requires exact equality between expected payoffs and aspiration levels for small trembles. In this respect it resembles Selten's (1975) notion of trembling hand perfect equilibrium, which approximates a pattern of behavior that is an *exact* Nash equilibrium for small trembles. This feature often complicates the computation and characterization of equilibria. We therefore present a notion of equilibrium that is slightly weaker than an ECA, wherein for sufficiently small trembles, aspiration levels should be arbitrarily near (instead of exactly equal to) the expected payoffs of a limit distribution of the Markov process induced by these aspirations.

An *equilibrium with nearly consistent aspirations* (ENCA) is a probability distribution μ^* over S and associated aspiration levels F^* and G^* with the property that there exist sequences of strictly positive tremble probabilities $\epsilon_k \to 0$, aspirations (F_k, G_k) and probability distributions μ_k over S such that

(1) μ_k is the limit distribution under (F_k, G_k, ϵ_k).

(2) μ_k converges (weakly) to μ^* and (F_k, G_k) converges to (F^*, G^*) as $\epsilon_k \to 0$.

(3) (F^*, G^*) is the expected payoff vector under the measure μ^*.

It can easily be verified that given the continuity property (A1), this is equivalent to requiring that the limit aspirations (F^*, G^*) along with the tremble sequence $\{\epsilon_k\}$ induce a Markov process with limit distribution $\mu(F^*, G^*, \epsilon_k)$ which converges to μ^*, and such that the corresponding sequence of expected payoffs converges to (F^*, G^*). The difference from ECA therefore lies in the property that the sequence of distributions μ_k is permitted to be induced by the *limit* aspirations, which are close but not exactly equal to the expected payoffs under μ_k.

It is evident from the definition that an ECA is also an ENCA, so Theorem 4.1 also ensures the existence of an ENCA. Indeed, under the continuity assumption (A1), one might argue that there is no need to worry excessively about aspirations exactly equalling long run average payoffs.

5 Preliminary Analysis: Tremble-Free Learning

It will be useful to start our analysis by considering the case where there are no trembles and aspirations are fixed. The properties of the learning process in this case will serve as stepping stones to the later analysis. Further, these results are of interest in their own right for a specific class of games, where the payoffs for each player can occupy only two possible values.[13]

We start by introducing some assumptions on learning rules which will be maintained throughout the paper. Roughly speaking, these assumptions embody the idea that actions that do well relative to aspirations are given more weight, and actions that do not are not exclusively adopted, as the player must shift some weight to other actions.

A player is said to receive a *nonzero feedback* (at some outcome) if the resulting payoff for that player differs from his aspiration level.

Fix outcome $c = (a, b)$. If player A's payoff exceeds his aspiration level F, the action a is said to generate *positive direct feedback* (PDF) under c and all other actions $a' \neq a$ are said to generate *negative indirect feedback* (NIF) under c. In the opposite case where A's payoff is less than his aspiration level, the above are replaced by the corresponding notions of *negative direct feedback* (NDF) and *positive indirect feedback* (PIF).

Of course, similar definitions apply to player B.

(A2: adaptation to positive direct feedbacks) *If an action receives PDF, it must be played with positive probability in the following period, and the probability weight on it at the next stage must increase at least at some positive geometric rate. Formally, suppose a is played with payoff f, which exceeds the aspiration level F.*

[13] Such contexts have been considered in the psychology literature dealing with stimulus reinforcement learning. The reason is that for such games the nature of the stimulus from any outcome is defined independently of a notion of an aspiration level, since obtaining the higher payoff must constitute a success, and the lower payoff a failure.

Then there exists $k \in (0,1)$*, possibly depending on* (f,F)*, such that* $0 < \alpha'(a) \geq \alpha(a)^k$ *for every* $\alpha \in \Delta(\mathcal{A})$ *(and every action* $a \in \mathcal{A}$*).*

Note that (A2) and (A.1)together imply that in the case of PDF, a must be played in the following period with a probability bounded away from zero in α, though this lower bound may depend on the realized payoff f or the aspiration level F.[14]

We now introduce the remaining assumptions.

(A3: adaptation to negative feedbacks) *If an action receives NDF, all other actions are played next period with positive probability.*

(A4: adaptation to other indirect feedbacks) *If an action receives one PIF followed by one NIF, then it is played with positive probability in the next period.*

(A5: slow adaptation) *If an action is selected with at least the average probability per action in any period, then it must be selected with positive probability in the following two periods.*

Using the same argument as in the case of (A2), assumption (A3) implies (in conjunction with (A1)) that the weight on any action receiving a NIF in the previous period is bounded away from zero. The bound is uniform over all ongoing strategies and actions, though it may depend on the achieved payoff and the aspiration level.[15]

Assumption (A4) "gives an unplayed action a chance" when it could have been played as the result of an earlier PIF (just as in (A3)) but wasn't, because a third action was played instead. The assumption requires that even if this third action receives PDF, the unplayed action must still be chosen with positive probability in the period after. As in the case of (A3), this probability is bounded away from zero, given (A1), with the bound being uniform with respect to the action in question or the ongoing strategy. Finally, (A5) requires that a single unfavorable experience with a currently "not unfavored" action does not cause its weight to fall below some positive bound in the following two periods. We reiterate that these probabilities may vary with the size of the feedbacks involved, though uniformity in the going mixed strategies is required.

Finally, we note that the Bush-Mosteller (1955) model of learning satisfies all these assumptions.[16]

---

[14] This follows from the fact that for fixed (f,F,a), a minimum of $\alpha'(a)$ with respect to choice of α over the compact set $\Delta(\mathcal{A})$ must be attained at some α^*, and (A2) ensures that this lower bound is strictly positive. After all, the assumption applies to *every* contingency where a is played, including those in which $\alpha(a) = 0$.

[15] Formally, fix f, and F with $f < F$. Suppose that a is played and receives payoff f. Then there exists $\pi \in (0,1)$, possibly depending on (f,F), such that for all $\alpha \in \Delta(\mathcal{A})$, a' is played next period (under $L^A(\alpha,a,f,F)$) with probability at least π, for every action $a' \neq a$.

[16] In the case of two actions, this model requires the current probability weight on an action to be replaced by a convex combination of the current weight, and one or zero, according as the action received a positive or negative feedback. The relative weight on the current probability can depend on the action chosen and the exact payoff realized. As long as this weight is always between zero and one, the reader can verify that all of the above assumptions are satisfied.

The main result of this section concerns the long run outcome of the game when the two players use learning rules satisfying properties (A1)-(A5), and the tremble probability is zero. We will need the following definitions. For given aspirations (F,G), define an outcome c to be *mutually favourable* (MF) if both players earn payoffs that strictly exceed their aspiration levels: $(f(c),g(c)) \gg (F,G)$. Next, for player A, define an action a to be *uniformly good (for A) and bad (for B)* (UGB) if choice of a results in a payoff for A which always exceeds player A's aspirations and a payoff for B which always falls short of B's aspirations, *no matter what action b* is chosen by B: $(f(a,b),-g(a,b)) \gg (F,-G)$ for all $b \in B$. A parallel definition holds for UGB actions for player B.

Say that there is an *infinite run* on some action vector if that action vector is played in every period barring at most a finite number of initial periods.

THEOREM **5.1** *Fix aspirations (F,G). Suppose that (A1)-(A5) hold, that neither player trembles with positive probability, that each player receives nonzero feedback from every outcome, and that there exists at least one MF action profile, or an action that is UGB for some player.*

Then from every initial measure on S, there must (almost surely) be an infinite run on a MF action pair, or on a UGB action by some player.

Proof: Denote by $\mathcal{E}$ the event that from the current date onward, some MF pair or some UGB action (for some player) is played forever, after the passage of at most 4 periods. It will suffice to prove that there exists $\epsilon > 0$ such that for *every* initial μ on S, the probability of the event $\mathcal{E}$ (conditional on μ) is at least ϵ, where ϵ is *independent* of μ. For then the event $\mathcal{E}$ must occur with probability one after some finite date, and this will establish the theorem.

Let $\delta > 0$ be the lower bound on the weight on an action which receives a PDF at the previous stage, as given by (A2) in conjunction with (A1). For given aspirations, this may depend on the achieved payoff level. But since there are only finitely many payoff levels, a positive lower bound on the weight can be constructed uniformly in the achieved payoff level also. Morever, without loss of generality, this bound applies to both players. Let S^* be the set of all states $\gamma = (\alpha,\beta)$ such that either $\alpha(a)\beta(b) \geq \delta^2$ for some MF (a,b), or $\alpha(a) \geq \delta$ for some UGB action a for A, or $\beta(b) \geq \delta$ for some UGB action b for B.

We prove the theorem in two steps. *In the first step*, we claim:

LEMMA **5.1** *There exists $\eta \in (0,1)$ such that starting from any initial distribution μ on S, the conditional probability of the state entering S^* in at most four periods is at least η.*

Proof of Lemma 5.1: We need the following notation. Recall that by (A3) (in conjunction with (A1)) that there is a lower bound $\pi > 0$ on the probability assigned to any action receiving a PIF at the previous stage. Just as for (A2), we can make this bound independent of payoffs as there are only finitely many of them. Likewise, we can find ϕ and θ both positive such that for any initial state $\gamma \in S$, and for either

player: (i) any action which receives a PIF in the first period, followed by a NIF, it will be played with probability at least ϕ in the subsequent period; (ii) any action played with at least average probability will be played in any of the subsequent two periods with at least probability θ.

Finally define $\eta = \frac{1}{|A|}\frac{1}{|B|}\delta^3\pi^4\theta\phi$. Note that $\eta > 0$.

Consider any initial measure μ on S. If $\mu(S^*) \geq \eta$, there is nothing to prove. Otherwise $\mu(S^*) < \eta$. In this case, observe that *some* action pair (a,b) is chosen with at least average probability.

Case I. *For some player (say A) there is a uniformly good (UG) action a^*.*

Consider an initial state γ, and an action pair $c = (a,b)$ such that c is selected with probability at least $\frac{1}{|A|}\frac{1}{|B|}$. If a is UGB there is nothing to prove; so suppose that a is not UGB. Consider first the case where $a = a^*$, i.e., it is UG but not UGB. This implies that there exists b^* such that (a^*,b^*) is MF. If $b = b^*$ as well, we are in S^* already. So suppose that b receives NDF under c. Then b^* will be played in the following period with probability at least π, while $a = a^*$ will be played with probability at least δ, as it received PDF in the current period. Hence with probability at least $\frac{1}{|A|}\frac{1}{|B|}\delta\pi$, (a^*,b^*) will be played in the following period, so that we shall be in S^*.

So now consider the case where $a \neq a^*$. If a is UG also, then since it is not UGB, a similar argument as above will establish the result. So we suppose that a is not UG. If a gets a NDF in c, then at the following period a^* will be selected with probability at least π; thereafter use the argument of the preceding paragragh to infer that in at least two periods we shall be in S^* with probability at least η. On the other hand if a gets PDF under c, there are two possibilities: c is MF, in which case there is nothing to prove, and a gets PDF while b gets NDF under c. Then a will be played in the following period with probability at least δ. Morever, since a is not UG, there exists $b' \neq b$ such that a receives NDF when b' is played. Then (a,b') will be played in the following period with probability at least $\frac{1}{|A|}\frac{1}{|B|}\delta\pi$. Conditional on this event, the period after that a^* will be played with probability at least π, and we shall be in S^* within four periods with probability at least η, again using the argument of the previous paragraph.

Case II. *There exists no UG action for either player, but there does exist an MF pair $c^* \equiv (a^*,b^*)$.* Let $c = (a,b)$ be played with probability at least $\frac{1}{|A|}\frac{1}{|B|}$ in the current period. If c is MF there is nothing to prove. This leaves the following possible subcases.

Subcase 2a: *Both players receive NDF under c.* Use the notation t for the current period. If $a \neq a^*$, $b \neq b^*$, then clearly c^* will be played at $t+1$ with probability at least π^2. If on the other hand $a = a^*$, $b \neq b^*$, then a will be played again at $t+1$ with probability at least θ by virtue of (A5). Morever b^* will be played with probability at least π at $t+1$. Consequently c^* shall be played with probability at least $\frac{1}{|A|}\frac{1}{|B|}\theta\pi > \eta$ at $t+1$.

Subcase 2b: *One player (say, A) gets a PDF, and the other player gets an NDF under c.* If for some $b^* \neq b$, (a,b^*) is MF, then this will be played at $t+1$ with

probability at least $\frac{1}{|\mathcal{A}|}\frac{1}{|\mathcal{B}|}\delta\pi > \eta$. *So suppose for the rest of this subcase that* $(a, \bar{b})$ *is not MF for any* $\bar{b}$.

Since a is not UG, there must exist $b' \neq b$ such that A gets NDF when B plays b'. Morever, (a, b') will be played at $t+1$ with probability at least $\frac{1}{|\mathcal{A}|}\frac{1}{|\mathcal{B}|}\delta\pi$. Following this event, there are two possibilities to consider:

(i) Suppose that Player B gets NDF under (a, b'). If $a \neq a^*$ and $b' \neq b^*$, then c^* will be played at $t+2$ with (conditional) probability at least π^2. On the other hand, if $a \neq a^*$, $b' = b^*$, note that a^* receives a PIF at $t+1$ when (a, b') is played. Then at $t+2$, (a, b) is played with (conditional) probability at least $\theta\pi$, since $b' \neq b$, and a was being played with at least average probability at $t+1$ as it received a PDF at t. Then a^* gets a NIF at $t+2$, while b gets a NDF. At $t+3$ then, $c^* = (a^*, b^*)$ will be played with conditional probability at least $\phi\pi$, and so we shall enter S^* in three steps with probability at least $\frac{1}{|\mathcal{A}|}\frac{1}{|\mathcal{B}|}\delta\theta\phi\pi^3 > \eta$.

(ii) Suppose that Player B receives PDF when (a, b') is played at $t+1$, following play of (a, b) at t. Recall that this sequence of play arises with probability at least $\frac{1}{|\mathcal{A}|}\frac{1}{|\mathcal{B}|}\delta\pi$. Recall, too, that $a \neq a^*$, since we are in the situation where $(a, \bar{b})$ is not MF for any $\bar{b}$.

Here there are three further subcases to consider:

(a) $b' = b^*$. Here a^* is played at $t+2$ with (conditional) probability at least π, and $b' = b^*$ with (conditional) probability at least δ. So the MF outcome results in period $t+2$ with probability at least $\frac{1}{|\mathcal{A}|}\frac{1}{|\mathcal{B}|}\delta^2\pi^2$.

(b) $b^* \neq b$, $b^* \neq b'$. In this case b^* obtained a PIF at t, and a NIF at $t+1$. So it is played at $t+2$ with probability at least ϕ, whilst a^* is played with probability at least π. So c^* is played at $t+2$ with probability at least $\frac{1}{|\mathcal{A}|}\frac{1}{|\mathcal{B}|}\delta\pi^2\phi$.

(c) $b^* = b$. In this case, note that by virtue of (A5), b^* is selected at $t+2$ with probability at least θ. Morever, since $a \neq a^*$ receives a NDF at $t+1$, a^* is played at $t+2$ with probability at least π. Hence c^* is played at $t+2$ with probability at least $\frac{1}{|\mathcal{A}|}\frac{1}{|\mathcal{B}|}\delta\pi^2\theta$. This concludes the proof of the Lemma. ■

We now turn to the second main step of the proof. Fix any MF pair or a UGB action for some player. We will denote by $\Pi(\gamma)$ the probability that starting from an initial state γ, the model generates an infinite run of the chosen pair (or action). Denote by $p(\gamma)$ the probability that the MF pair (or UGB action) is realized once. For example, if $\gamma = (\alpha, \beta)$ and (a, b) is MF, then $p(\gamma) = \alpha(a)\beta(b)$.

We claim that there exists $M \in (0, \infty)$ such that

$$\Pi(\gamma) \geq p(\gamma)^M \tag{7}$$

for all $\gamma \in S$.

Let $\Pi^t(\gamma)$ denote the probability of a run of length t (of the desired MF pair or UGB action). It is obvious that $\Pi^t(\gamma) \geq \Pi^{t+1}(\gamma)$ for each γ, and that $\Pi(\gamma)$ is just the pointwise limit of $\Pi^t(\gamma)$. It will therefore suffice to establish that there exists $M \in (0, \infty)$ such that

$$\Pi^t(\gamma) \geq p(\gamma)^M \tag{8}$$

for all $\gamma \in S$ and positive integers t.

Recall the definition of the rate of growth of weight on an action receiving a PDF, which may depend on the achieved payoff and the aspiration level: for player A let this be denoted $k^A(f,F)$ (and its counterpart $k^B(g,G)$) for player B). Let k be the maximum value of all $k(f(c),F)$ (if $f(c) > F$) and all $k(g(c),G)$ (if $g(c) > G$). By the conditions of the theorem, k is well defined and strictly between zero and one. Let M equal $(1-k)^{-1}$. Note that $M > 1$, so that

$$\Pi^1(\gamma) = p(\gamma) \geq p(\gamma)^M.$$

Now proceed by induction. Suppose that at time t, $\Pi^t(\hat{\gamma}) \geq p(\hat{\gamma})^M$ for all $\hat{\gamma} \in S$. Letting γ' denote the state following γ and the occurrence of the chosen outcome, observe that

$$\begin{aligned} \Pi^{t+1}(\gamma) &= p(\gamma)\Pi^t(p(\gamma')) \\ &\geq p(\gamma)[p(\gamma')^M] \\ &\geq p(\gamma)[p(\gamma)]^{kM} \\ &= p(\gamma)^M, \end{aligned}$$

where the first inequality is just the induction hypothesis, the second follows from applying (A2).

This completes the verification of the claim.

Finally, combining with Lemma 5.1, we obtain the property described in the first paragraph of the proof. Define $\epsilon \equiv \eta\delta^{2M}$. Then from every initial μ on S, the conditional probability of the event $\mathcal{E}$ occuring is at least ϵ. Therefore $\mathcal{E}$ must occur with probability one, and the proof is complete. ■

Note that if a game has no UGB actions, but it does have a MF outcome for given aspirations, then this theorem implies that play must eventually involve the selection of pure strategies by either player. In other words, limiting play must be degenerate. In the case where a UGB action is present, this is also true for at least one player: *e.g.* following repeated selection of a pure strategy corresponding to an UGB action by player A, player B will not be "satisfied" with any action, and may therefore be unable to settle down to a pure strategy. Note, however, that in the presence of multiple MF/UGB outcomes, the theorem merely says that infinite repetition of *one* of these is eventually inevitable, but it does not say which one will be selected.[17]

The theorem has particular implications for games with $0-1$ payoffs, where a payoff of 1 corresponds to a PDF for the chosen action, and a payoff of 0 to an NDF. Essentially, the notion of an aspiration is irrelevant in this context, since the notion of success or failure is unambiguous. In such a game, an MF strategy pair is one where both players earn a payoff of 1. A UGB strategy for a player is one

[17] In fact, starting from any initial state involving completely mixed strategies for both players, any given MF or UGB outcome will arise with positive probability.

in which that player earns 1 and the other player earns 0, regardless of the latter's strategy; in particular it must be a dominant strategy for the former. Hence, the theorem implies that if these exist, then there must be convergence either to a pure strategy Nash equilibrium, or to the selection of a dominant strategy by at least one player. If there are no UGB strategies in the game, but MF pairs exist, then we obtain convergence to a pure strategy Nash equilibrium for both players.

What about $0-1$ games without MF or UGB strategies? In such games, play will typically converge to a pattern of behavior where non-Nash strategies are selected with positive probability. In fact, even dominated strategies may survive in the long run.[18] This can happen because dominant strategies may occasionally generate failures, depending on what the other player happens to choose. This will induce players to perpetually experiment with other, possibly dominated, strategies.

6 Cooperation under ECA and ENCA: Some General Results

In this section we examine the extent of cooperation that can arise in an ECA or ENCA. Our first result pertains to *symmetric games*, defined as games in which both players have the same set of pure strategies ($\mathcal{A} = \mathcal{B}$), the same payoff functions ($f(a,b) = g(b,a)$ for any outcome $(a,b) \in \mathcal{C}$), the same learning rules ($L^A = L^B$) and the same trembles ($\nu^A = \nu^B$).

The *feasible payoff set* of a game is the convex hull of the set of pure strategy payoff vectors of the game.

THEOREM **6.1** *Assume (A1)-(A5). Consider a symmetric game and let* (π^*, π^*) *be a symmetric Pareto efficient pure action payoff. Suppose that* (π^*, π^*) *Pareto-dominates some other other feasible payoff vector. Then there exists an ECA* (μ^*, F^*, G^*) *such that* $(F^*, G^*) = (\pi^*, \pi^*)$ *and* μ^* *assigns probability one to the pure strategy profiles generating* (π^*, π^*)*, i.e. both players earn a payoff of* π^* *for sure.*

Proof. Denote by $\bar{\pi}$ the (common) expected payoff to each player when both players tremble. Then since trembles have full support, the assumptions of the theorem imply that $(\bar{\pi}, \bar{\pi})$ must be Pareto dominated by (π^*, π^*), *i.e.* $\bar{\pi} < \pi^*$.

Fix a half-open interval in $\mathbb{R}_{++}$ of the form $\mathcal{I} \equiv [\hat{\pi}, \pi^*)$ with the properties that (a) $\hat{\pi} > \bar{\pi}$, and (b) for each $\pi \in \mathcal{I}$, there is no outcome c with either $f(c) = \pi$ or $g(c) = \pi$. As there are only a finite number of distinct payoffs, such an interval must exist.

Consider symmetric aspirations of the form (F, F). We claim that if $F \in \mathcal{I}$, then (i) there is no UGB action for any player, and (ii) (π^*, π^*) is the *only* MF payoff. Part (i) of the claim follows from symmetry. If a is UGB for A, then a must also be UGB for B. But then at (a,a), both players must receive more than their aspirations, a contradiction to the assumption that a is UGB for A. To verify part (ii), note that

[18] Examples of these may be found in earlier versions of this paper, and are available on request.

if there is another MF payoff vector relative to (F, F), then by construction of the set $\mathcal{I}$, both players must be getting a payoff at least as high as π^*. So another MF payoff must Pareto dominate (π^*, π^*), contradicting the premise that the latter is Pareto efficient.

By Theorem 3.1, there exists a unique limit distribution $\mu^*(F, F, \epsilon)$ for each $F \in \mathcal{I}$ and tremble probability ϵ. Because the game is symmetric, because aspirations are symmetric, and because $\mu^*(F, F, \epsilon)$ is unique, $\mu^*(F, F, \epsilon)$ must also be symmetric.

Since (π^*, π^*) is the only MF payoff, we see by Theorem 5.1 that the only invariant measure for the system with $\epsilon = 0$ is a measure which places probability one on pure strategy profiles that generate (π^*, π^*). By Lemma 4.1, $\mu^*(F_n, F_n, \epsilon_n)$ converges to precisely this measure for any sequence $(F_n, F_n, \epsilon_n) \to (F, F, 0)$.

Let $e(F, \epsilon)$ be the (common) expected payoff generated by $\mu^*(F, F, \epsilon)$. The observations in the previous paragraph and the definition of $\bar{\pi}$ imply that $e(F, 1) = \bar{\pi}$ and $\lim_{\epsilon \to 0} e(F, \epsilon) = \pi^*$, for each $F \in \mathcal{I}$. Lemma 4.1 implies that $e(F, .)$ is continuous; hence, by the intermediate value theorem, for any $F \in \mathcal{I}$, there exists $\epsilon > 0$ such that $e(F, \epsilon) = F$.

Take a sequence F_n in $\mathcal{I}$ converging to π^*. Then there exists an asssociated sequence $\epsilon_n > 0$ satisfying $e(F_n, \epsilon_n) = F_n$. Take a subsequence ϵ_{n_k} converging to ϵ^*, say. We then have by Lemma 4.1: $\lim_k e(F_{n_k}, \epsilon_{n_k}) = e(\pi^*, \epsilon^*)$. But this limit must also be equal to π^* by construction. Hence $e(\pi^*, \epsilon^*) = \pi^*$. This implies that $\epsilon^* > 0$, since otherwise (π^*, π^*) must be in the interior of the feasible payoff set. We can then select the subsequence of aspirations (F_{n_k}, F_{n_k}), and strictly positive tremble probabilities ϵ_{n_k} converging to (π^*, π^*) and 0 respectively, to establish the result. ■

A number of well-known games satisfy the conditions of this theorem: e.g. the Prisoners' Dilemma, symmetric coordination games with multiple Nash equilibria, as well as symmetric Cournot duopolies. It implies, for instance, the existence of an ECA in the Prisoners' Dilemma in which both players cooperate with probability one, and of an ECA in a symmetric coordination game in which players successfully coordinate on the Pareto-optimal Nash equilibrium. It implies that symmetric Cournot duopolists can achieve collusive outcomes without any explicit coordination and with decision-making processes involving modest degrees of rationality, as long as they have appropriate self-justifying aspirations.

The idea behind this result is the following. If players aspire to cooperative payoff levels (strictly speaking, slightly less than these levels, to allow for infrequent trembles and deviations), then the cooperative outcome is the only mutually favourable one. Since the game is symmetric and so are aspirations, neither player can have a UGB strategy. Theorem 5.1 thus ensures that when players do not tremble at all, there is a unique limiting outcome where they must eventually cooperate all the time. Hence this must also be approximately true when players tremble slightly, so they must end up with average payoffs which are indeed close to the cooperative level, thereby justifying their initial aspirations.[19]

[19] By the assumptions of the theorem, achieved payoffs for both players will indeed be strictly less than the cooperative payoff when there are slight trembles.

Nevertheless, the proof is somewhat complicated by the requirement that aspirations and average payoffs must be *exactly* equal for small trembles, *i.e.*, by the need to verify the existence of a sequence of fixed points in the neighbourhood of zero trembles. This is where the symmetry assumption was useful. Without this assumption, we establish below an analogous result by weakening the solution concept to ENCA. Whether such a weakening is necessary remains an open question.

THEOREM **6.2** *Assume (A1)-(A5).Suppose that (a,b) is a Pareto efficient pure strategy pair with associated payoff vector (F^*,G^*) that gives each player higher than his maxmin payoff. Then there exists an ENCA in which players attain (F^*,G^*) with probability one.*

Proof. For sufficiently small $\nu > 0$, the set B of payoffs (f,g) satisfying $F^* - \nu < f < F^*$ and $G^* - \nu < g < G^*$ has the property that for any aspiration vector in B, (F^*,G^*) is the unique MF outcome, while there are no UGB actions for either player (since both players get more than their maxmin payoff at (F^*,G^*)).

Take a sequence of aspiration vectors (F^r,G^r) in B converging to (F^*,G^*). For each r, select a sequence of positive trembles ϵ^{rv} converging to zero. Theorem 3.1 ensures that for each r, the corresponding sequence of limit distributions μ^{rv} is well defined. Since S is compact, this has a weakly convergent subsequence. Without loss of generality, restrict attention to this subsequence hereafter. When there are no trembles, Theorem 5.1 tells us that the long run payoff is unique and given by (F^*,G^*). It follows that for each r, the sequence of long run average payoffs $(\bar{F}^{rv},\bar{G}^{rv})$ generated under the trembled process converges to (F^*,G^*).

Construct the sequence $\mu^r = \mu^{rr}$, $\epsilon^r = \epsilon^{rr}$. Then $\epsilon^r \to 0$ and μ^r converges weakly to a distribution concentrated on the payoff (F^*,G^*). Combining with the aspiration sequence (F^r,G^r), we obtain an ENCA with the required property. ∎

The idea behind the proof is similar to that of the preceding theorem: if both players have aspirations slightly lower than the desired cooperative payoffs, then cooperation is the unique MF outcome, and must therefore be the sole limiting outcome, absent any trembles. Hence if the sequence of trembles converges to zero, the corresponding sequence of average payoffs must converge to the cooperative payoff vector. So an arbitrary sequence of aspiration levels converging to the cooperative payoff from "below" generates a corresponding sequence of induced average payoffs which also converges to the same payoff.

The preceding theorems justify the idea that cooperative outcomes are compatible with a long run equilibrium when players employ adaptive learning rules driven by self-justifying aspirations. The next obvious question pertains to the possibility of inefficient outcomes. The next result rules out the possibility of a large set of inefficient non-pure-strategy payoff outcomes.

THEOREM **6.3** *Assume (A1)—(A5). Consider any payoff vector (F^*,G^*) in the feasible payoff set of a game which gives neither player a pure strategy payoff, and is strictly Pareto-dominated by some pure strategy payoff vector. Then there cannot be an ENCA resulting in average payoffs (F^*,G^*).*

Proof. If there exists an ENCA with aspirations (F^*, G^*), there must exist: (i) a distribution μ^* on S such that (F^*, G^*) is the expected payoff vector under μ^*, (ii) a sequence of aspirations $(F^n, G^n) \to (F^*, G^*)$, and a sequence of positive trembles $\epsilon^n \to 0$, which generate (iii) a sequence of limiting distributions μ^n converging weakly to μ. Hence Lemma 4.1 implies that μ^* is an invariant distribution for aspirations (F^*, G^*) and zero trembles.

Consider the set of invariant distributions on S induced by aspirations (F^*, G^*) and zero trembles. Since (F^*, G^*) is Pareto dominated by a pure strategy pair, there exists at least one MF outcome relative to these aspirations. Since neither player receives a pure strategy payoff in (F^*, G^*), feedback effects are always non-zero in the game with aspirations fixed at (F^*, G^*). Morever, given any set of aspirations, there can be at most one player who has a UGB strategy, by definition. Without loss of generality, let this be player A. Every invariant distribution corresponding to aspirations (F^*, G^*) and zero trembles must then mix the different MF and UGB outcomes relative to these aspirations, by virtue of Theorem 5.1. It follows that in every such invariant distribution, player A must obtain an average payoff strictly exceeding F^*. This contradicts the conclusion of the previous paragraph. ■

In the case of a Prisoners' Dilemma, this result rules out (almost) all inefficient payoffs except that of mutual-defection, as well as a range of asymmetric payoffs where one player gets more than the mutual-cooperation payoff while the other player obtains less (see Figure 1).

Insert Figure 1

In the case of a game with *common payoffs*, i.e., where both players obtain equal payoffs in every outcome, the theorem rules out all mixed strategy payoffs. Hence, in a game of this kind, only a finite number of payoff vectors are sustainable as an ENCA, i.e., only pure strategy payoffs. This contrasts sharply with the "large" number of outcomes sustainable as Nash or subgame perfect equilibria, as manifested in the well-known Folk Theorems.

The question still remains: could an ECA or ENCA result in a Pareto-inefficient outcome? The preceding theorem suggests that we should examine this with particular care for inefficient *pure* action outcomes. For instance, could mutual defection be a long run outcome in the Prisoners' Dilemma? Can a Pareto-inefficient pure strategy Nash equilibrium be an ECA? These issues are addressed in the following section.

7 Some Examples

In this section, we consider examples of two specific games: the Prisoners' Dilemma and a coordination game. The objective is partly to explore the tractability of our approach when applied to specific games, and partly to examine the questions raised at the end of the previous section.

To simplify the exposition, we assume in this section that players choose probability weights from an arbitrary finite grid Q on the interval $[0,1]$, which includes both 0 and 1. The state space is then the set of all mixed strategy pairs (α,β) with the property that α and β assign probabilities in the grid Q to every action. The state space is therefore discrete, and for given aspirations the learning rules of the two players generate a finite Markov chain. This implies that we can invoke the results concerning invariant distributions of perturbed finite Markov chains presented in Young (1993, Appendix). We shall hereafter refer to such contexts as involving *discrete learning rules.*

In this framework Assumption (A1) reduces to requiring continuity of learning rules in the aspiration level, while (A2)–(A5) can be translated appropriately. While the translation of (A3)-(A5) is straightforward, (A2) would imply here that a PDF on an action must result in an increase in the weight on that action in the following period (unless the weight already happens to be 1). This increase in weight, however, must consist of at least a certain discrete amount. Consequently, as it stands, assumption (A2) could well be inconsistent with (A1): vanishingly small PDFs must result in a least a discrete increase in the weight on an action, whereas a zero feedback will typically cause the current strategy to be unchanged. To restore continuity of the learning rule in aspirations, we therefore modify the rule to make it non-deterministic in the following fashion.

With a certain *inertial* probability q, a player leaves her current strategy unaltered, whilst with the residual probability $1-q$, she modifies it in a deterministic fashion satisfying (A2)-(A5). We impose the following assumptions on the inertial probability:

(A6: Inertia) *The inertial probability q is a continuous function of the achieved payoff f and the aspiration level F, bounded away from zero, and satisfies $q=1$ if the feedback is exactly zero ($f=F$).*

For the rest of this section we shall assume that the learning rule satisfies (A1)-(A6). It can be verified that all previous results of this paper continue to hold in this context.[20] Nevertheless, to provide insight into the nature of our results, we shall rarely invoke any of our earlier theorems, and will provide direct arguments instead; indeed, these will also serve to provide illustrations of the logic of the earlier results. In keeping with the idea that these examples serve expositional purposes, we suppress tedious technical details in the proofs.

We shall consider the following class of games which includes both a coordination game and the Prisoners' Dilemma as special cases.

	C	D
C	(b,b)	(0,x)
D	(x,0)	(a,a)

[20] This is shown in earlier versions of this paper; a proof is available on request.

where $b > a > 0$. If $x < b$ then this reduces to a coordination game, with two pure strategy Nash equilibria: (C,C) and (D,D), with the former Pareto-dominating the latter. If on the other hand $x > b$ this is a Prisoners' Dilemma. For simplicity, we ignore the borderline cases where x may equal a or b.

We now present a basic set of results concerning possible ENCAs of this class of games. These will be combined later to yield conclusions concerning the coordination game or the Prisoners' Dilemma.

THEOREM **7.1** *Consider the class of games, represented above in tabular form, with $x \neq a$, $x \neq b$. Assume that players employ discrete learning rules satisfying (A1)-(A6). Then*

(i) There always exists an ECA resulting in the cooperative payoff (b, b).

(ii) There exists an ENCA resulting in the payoff (a, a) if and only if $x > a$.

Proof. (i) Consider the invariant distribution corresponding to the case where both players have aspirations equal to b, and neither of them trembles. Then it is easily verified that the pure strategy state where both players select C for sure, is the unique absorbing state, and morever can be reached in a finite number of steps from any other state with positive probability (Figure 2 below illustrates the dynamic induced separately for the two cases where $x < b$ and $x > b$ respectively).

Insert Figure 2

Hence with zero trembles there is a unique invariant distribution which is concentrated entirely on this outcome. So the sequence of limiting distributions corresponding to any sequence of asymptotically vanishing trembles and aspirations (b, b) must converge to this degenerate distribution, where the payoffs equal (b, b).

This argument proves that (C, C) is ENCA; it is an ECA by an argument analogous to Theorem 5.1.

(ii) We first show that if $x > a$ then there exists an ENCA resulting in the payoffs (a, a). The dynamic pattern is shown in the first part of Figure 3. With $x > a$, the D strategy becomes UG for each player.

Insert Figure 3

There are two absorbing states, corresponding to the pure strategy outcomes (C,C) and (D,D), while all other states are transient. So every invariant distribution must concentrate its weight on either of these. Note also that one tremble is required to move from the absorbing state (C,C) to the other (D,D), but two trembles are required in the opposite direction. In the terminology of Young (1993), the (D,D) state has a lower stochastic potential than (C,C). So the limit of the sequence of (unique) invariant distributions corresponding to an arbitrary sequence of vanishing trembles must be the invariant ditribution concentrated entirely on (D,D), where both players earn a payoff of a.

Now consider the case where $x < a$, so that D is no longer a UG strategy. The resulting dynamic pattern is shown in the second part of Figure 4. It is still the case

that the pure strategy states (C,C) and (D,D) are the only absorbing states, and all other states are transient. The only difference is in the pattern of resistances: now one tremble suffices to go from any one of these two absorbing states to the other, so they have the same stochastic potential. Specifically, starting from the pure (D,D) state, one tremble (by A, say) serves to move the state to one where A selects C with positive probability, whilst B continues to select the pure strategy D. Hence the outcome (C,D) occurs with positive probability. Given $x < a$, B will now assign positive weight to C by virtue of (A3), whilst (A6) implies that with positive probability A will not change his current strategy owing to inertia. Hence with positive probability the state will transit to one where both select C wth positive probability, and thence to the one where they both select C with probability one. Hence one tremble suffices to move from the absorbing state (D,D) to the other (C,C). The same is true in the opposite direction. So the limit of the invariant distributions corresponding to vanishing trembles must assign positive weight to both outcomes (C,C) and (D,D), implying that the limiting payoffs must strictly exceed a. Since the learning rule is continuous, this establishes that there cannot be an ENCA resulting in payoffs (a, a). ∎

We are now in a position to provide a more complete characterization of equilibria in coordination games, where $x < b$.

THEOREM **7.2** *Consider a generic coordination game, where $x < b, x \neq a$, and suppose that both players employ discrete learning rules satisfying (A1)-(A6). Then:*

(i) If $x < a$, there is a unique ENCA (and ECA) outcome where both players earn a payoff of b.

(ii) If $x > a$, there are two ENCA payoff vectors: (b, b) and (a, a).

Proof. Applying an argument analogous to that of Theorem 5.3, no mixed strategy payoff can be the result of an ENCA, as it is Pareto dominated by the pure strategy payoff (b, b). Hence the only possible ENCA outcomes are the pure strategy payoffs $(b, b), (a, a), (x, 0), (0, x)$. Given Theorems 5.1 and 6.1, it remains to show that $(x, 0)$ or $(0, x)$ can never be an ENCA outcome. Given (A1), it suffices to check that with aspirations equal to $(x, 0)$, the corresponding limit of the invariant distributions (as trembles go to zero) yields an average payoff different from $(x, 0)$; an analogous argument takes care of $(0, x)$. The dynamics yielded by aspiration vector $(x, 0)$ in the case of no trembles is illustrated in Figure 4 below, for both the cases $x < a$ and $x > a$.

Insert Figure 4

In the case where $a > x$, note there are three absorbing states: those corresponding to the pure strategies (C,C), (D,D) and (D,C). All others are transient. It takes two trembles to move from the pure strategy state (D,D) to the pure strategy state (C,C), whereas it takes exactly one tremble to move between any other pair of absorbing states. Hence the limiting invariant distribution (as trembles vanish) must divide weight between the two pure strategy states (D, C) and (D, D). Then

both players earn higher than their aspirations on average, so $(x,0)$ cannot be an ENCA outcome in this case. In the other case, it is evident that there are only two absorbing states: the pure strategy states corresponding to (D,C) and (C,C), while all others are transient. Morever, one mistake suffices to move from one absorbing state to the other, so the average payoff in the limiting invariant distribution must be higher for both players than their aspirations. ■

The coordination game thus exhibits the following features: (a) long run equilibria must involve one of the two pure strategy Nash equilibrium payoffs; (b) for some parameter values ($x < a$), the long run equilibrium outcome is unique and efficient; (c) for other parameter values, there are multiple Pareto-ranked equilibria, so in this range initial conditions will matter in determining the eventual outcome. Note in particular the contrast with the theories of equilibrium selection of Kandori, Mailath and Rob (1993) or Young (1993): long run outcomes need not be unique, and the inefficient Nash outcome can result even if it is both Pareto-dominated and risk-dominated by the efficient Nash outcome.[21]

Finally, we consider the Prisoners' Dilemma where $2b > x > b > a > 0$. Using the result of Theorem 6.1, the following is immediate.

THEOREM **7.3** *Consider the Prisoners Dilemma, with $x > b > a > 0$ and suppose both players employ discrete learning rules satisfying (A1)-(A6). Then the mutual cooperation payoff is an ECA, and the mutual defection payoff is ENCA.*

Note the contrast with the coordination game, where the inefficient pure strategy outcome of mutual-D could not be an ENCA. The reason is the difference in the "off-equilibrium" dynamic: the outcome CD leads both players in the direction of mutual defection in the Prisoners' Dilemma, whilst in the coordination game it causes the two players to move in opposite directions.

Indeed, in the Prisoners' Dilemma, mutual-D is even an ECA under additional restrictions. A proof of this assertion is available on request.

There is one other distinct feature of the Prisoners' Dilemma: (C,C) does not Pareto dominate all other strategy pairs. In particular, there are asymmetric mixed strategy outcomes that are Pareto efficient and individually rational (e.g., alternation between (C,C) and (C,D) in suitable proportions). This raises the possibility of long run outcomes apart from (C,C) or (D,D) involving asymmetric payoffs. What can be shown is the following: (a) an ENCA outcome with symmetric payoffs must be either of the two pure strategy outcomes (C,C) or (D,D), and (b) an ENCA outcome with asymmetric payoffs must involve one player obtaining an expected payoff exceeding b, and the other player between a and b. Part (a) follows from the fact that the long run outcome of the process induced by aspirations for both players intermediate between a and b is (C,C) with zero trembles. Part (b) follows by noting that no player can get less than a, his maxmin level, while the other player gets more than a — essentially because with these aspirations D is a UGB strategy for the former. If the other player has an aspiration exceeding b then there are no other MF or

[21] To illustrate this last point, consider the case where $b > x + a$ and $x > a$.

UGB outcomes, so the low aspiration player must end up with payoffs of at least a.[22] Hence an asymmetric outcome can only take the form of one player defecting "more" than the other on average, while they cycle between different states.

8 Conclusion

Non-Nash play can arise in a game for two reasons. First, the learning rule is sometimes incapable of finding the best outcome *even in one-person "games" against a deterministic environment.* Second, the intrinsic structure of interactive play may prevent each player from choosing a best response to the other's strategy. It should be emphasized that in our model, the *second* feature drives our results. In the one-person variant of this model (in a deterministic environment) an equilibrium with consistent aspirations always yields the highest possible payoff.

Nevertheless, our results must be qualified by the following observations.

(1) In games with many players, our results are weakened on two counts. First, in the event of a defection the coordination required to restore cooperation is of an order of magnitude that is exponential in the number of players. In addition, any single player's deviation has a smaller effect, thereby dampening the reactions of the opponents. In particular, cooperation is impossible with a continuum of players. Limit play must be Nash.

(2) In "learning" games (even with a few players) where the past is given the same weight as the present, but where play is otherwise myopic (fictitious play, for instance), the tendency to cooperate is also weakened. A current deviation does not evoke a large reaction from the opponent, who considers her rival's entire history of play.

(3) Our results therefore pertain to environments where there are a small number of players *and* where current experience is paramount in determining current strategy. There is also the somewhat more nebulous issue of *how much rationality* is assumed for the players. Following the spirit of an evolutionary perspective, one might well ask: what degree of rationality is conducive to cooperation? Assuming that players behave myopically, *i.e.* try to increase current payoffs, the answer might be a "low degree of rationality" in some well-defined sense.[23]

Numerous avenues of future research appear promising. First, it is essential to develop a more complete theory by explicitly modelling the dynamics of aspiration

[22] The same outcome results even if the other player has an aspiration intermediate between a and b, whence (C,C) is MF but also one where the low aspiration player gets a payoff higher than a.

[23] As Vernon Smith (1991) elaborates, game experiments where players are better informed (such as in experimental asset markets, rather than double auctions) usually exhibit slower convergence to efficient equilibrium outcomes. Morever, that play eventually does tend to converge to these equilibria stands in contrast to persistent deviations from "rational" behaviour exhibited in many single person experimental contexts. Our approach posits behaviour at the individual decision-maker's level which is consistent with observed deviations from rational play in many single person experiments (such as emergence of "matching" behavior in two arm bandit problems), while showing that cooperative behavior may nevertheless obtain in a multiperson setting.

revision. Second, the sensitivity of our results to alternative formulations of satisficing ought to be explored. Third, the model should be extended to accommodate more than two players. Fourth, it would be interesting to examine the consequences of embedding this model within a richer setting, *e.g.* where firms select prices, customers select products, voters select between different candidates and so on.

References

Arthur, B. (1993), "On Designing Economic Agents that Behave like Human Agents," *Journal of Evolutionary Economics* **3**, 1–22.

Binmore, K. and L. Samuelson (1993), "Muddling Through: Noisy Equilibrium Selection," mimeo, University College, London.

Bush, R. and Mosteller, F. (1955), *Stochastic Models of Learning.* New York: John Wiley and Sons.

Canning, D. (1989), "Convergence to Equilibrium in a Sequence of Games with Learning," ICERD Discussion Papert, London School of Economics.

———- (1991), "Average Behavior in Learning Models," *Journal of Economic Theory*, 57:442-472.

Estes, W. (1954), "Individual Behavior in Uncertain Situations: An Interpretation in Terms of Statistical Association Theory," in *Decision Processes*, edited by R.M. Thrall, C.H. Coombs and R.L. Davis, New York: Wiley.

Feller, W. (1957), *An Introduction to Probability Theory and its Applications.* New York: John Wiley and Sons.

Gilboa, I. and D. Schmeidler (1992), "Case-Based Decision Theory," mimeo, Kellogg School of Management, Northwestern University.

Fudenberg, D. and Kreps, D. (1988), "A Theory of Learning, Experimentation and Equilibrium in Games," mimeo, Department of Economics, MIT.

Jordan, J. (1991), "Bayesian Learning in Normal Form Games," *Games and Economic Behavior*, 3(1), 60-81.

Kandori, M., G. Mailath and R. Rob (1993), "Learning, Mutation and Long Run Equilibria in Games," *Econometrica.*

Kohlberg, E. and J. F. Mertens (1986), "On the Strategic Stability of Equilibrium," *Econometrica.*

Krishna, V. (1992), "Learning in Games with Strategic Complementarities, " mimeo, Harvard Business School.

March, J. and H. Simon (1958), *Organizations*, New York, Wiley.

Milgrom, P. and Roberts, J. (1991), "Adaptive and Sophisticated Learning in Normal Form Games," *Games and Economic Behavior*, 3(1), 82-100.

Mookherjee, D. and Sopher, B. (1992), "Learning Behaviour in an Experimental Matching Pennies Game," forthcoming, *Games and Economic Behavior.*

———————- (1993) "Learning and Decision Costs in Constant Sum Experimental Games, " mimeo, Rutgers University.

Nelson, R. and Winter, S. (1982). *An Evolutionary Theory of Economic Change.* Cambridge, Massachussetts: Harvard University Press.

Robinson, J. (1951), "An Iterative Method of Solving a Game," *Annals of Mathematics*, 54, 296-301.

Roth, A. and I. Erev (1993), "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," mimeo, University of Pittsburgh.

Selten, R. (1975), "A Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25-55.

———- (1978), "The Chain Store Paradox," *Theory and Decision.*

———- (1991), "Evolution, Learning and Economic Behavior," *Games and Economic Behavior*, 3(1).

——— and Stoecker, R. (1986), "End Behavior in Sequences of Finite Prisoners' Dilemma Supergames," *Journal of Economic Behavior and Organization* , 7, 47-70.

Simon, H. (1955), "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69, 99-118.

———- (1957), *Models of Man*, New York.

———- (1959), "Theories of Decision Making in Economics and Behavioral Science," *American Economic Review*, 49(1), 253-283.

Shapley, L.S. (1964), "Some Topics in Two Person Games," in *Advances in Game Theory, Annals of Mathematical Studies*, 5, 1-28.

Smith, V.L. (1991), "Rational Choice: The Contrast Between Economics and Psychology," *Journal of Political Economy*, 99(4), 877-898.

Suppes, P. and Atkinson, R. (1960). *Markov Learning Models for Multiperson Interactions,* Stanford University Press, Stanford.

Van Damme, E. (1987). *Stability and Perfection of Nash Equilibria,* Springer-Verlag, Berlin.

Winter, S. (1971), "Satisficing, Selection and the Innovating Remnant," *Quarterly Journal of Economics*, 237-260.

Young P. (1993), "The Evolution of Conventions," *Econometrica*.

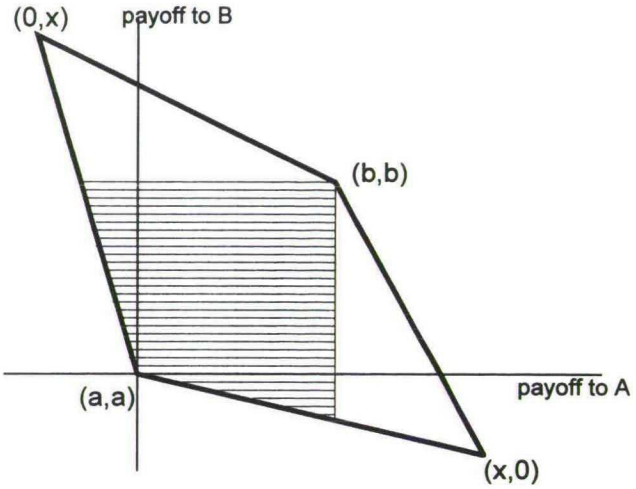

Figure 1: Prisoners' Dilemma. Shaded area (except for (a,a)) ruled out.

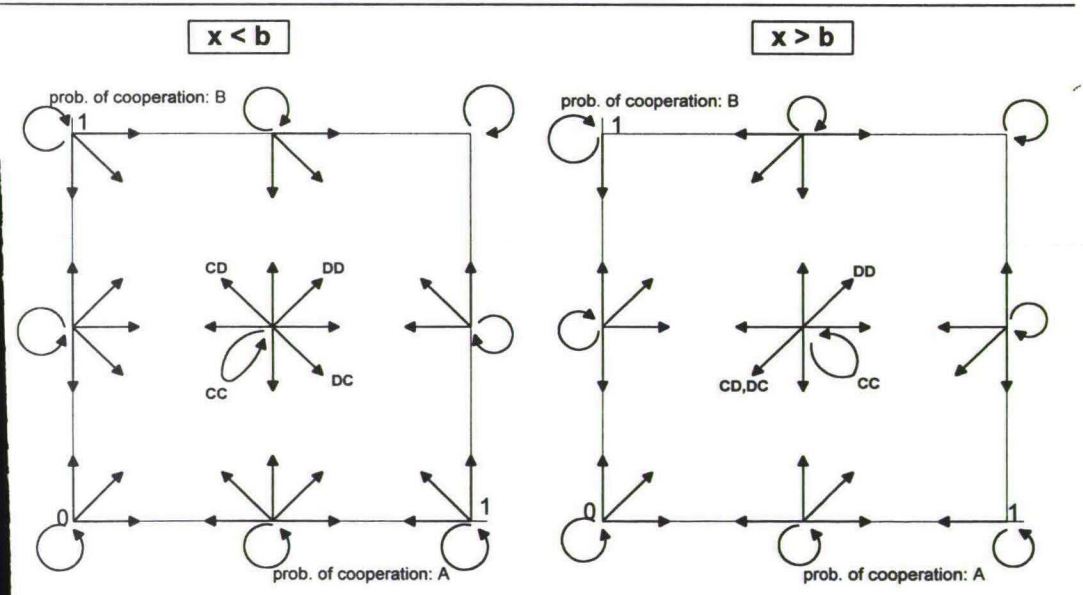

Figure 1: Diagrams for the proof of Theorem 7.1, part (i).

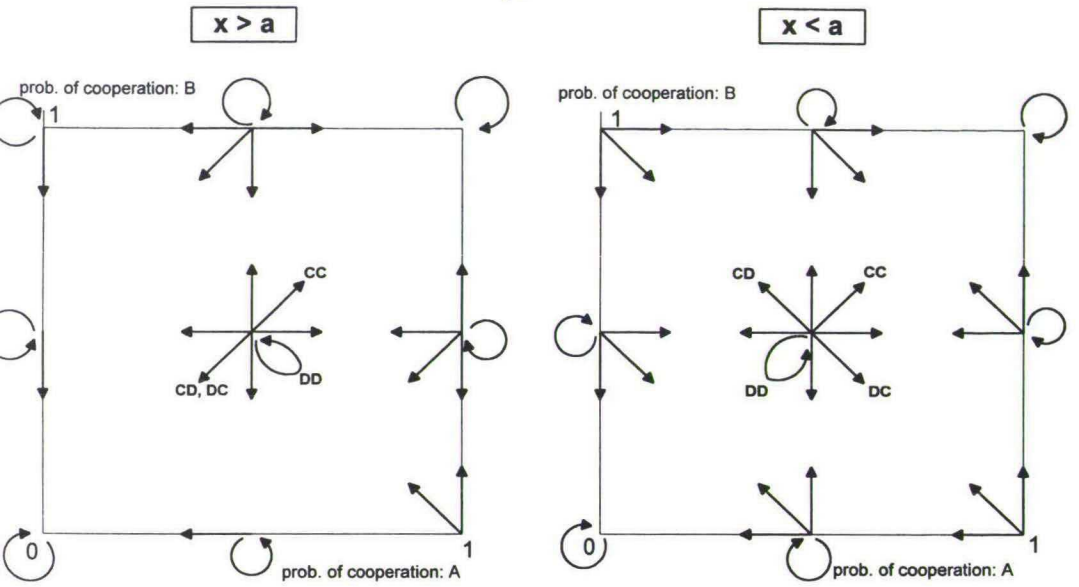

Figure 3: Diagrams for the proof of Theorem 7.1, part (ii).

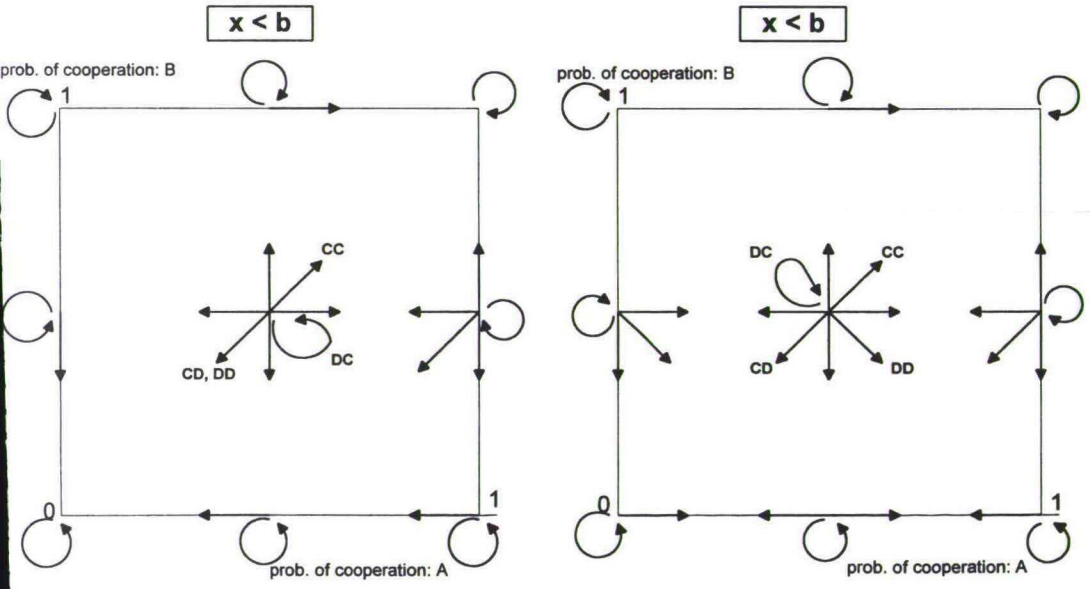

Figure 4: Diagrams for the proof of Theorem 7.2.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9335	A.L. Bovenberg and F. van der Ploeg	Green Policies and Public Finance in a Small Open Economy
9336	E. Schaling	On the Economic Independence of the Central Bank and the Persistence of Inflation
9337	G.-J. Otten	Characterizations of a Game Theoretical Cost Allocation Method
9338	M. Gradstein	Provision of Public Goods With Incomplete Information: Decentralization vs. Central Planning
9339	W. Güth and H. Kliemt	Competition or Co-operation
9340	T.C. To	Export Subsidies and Oligopoly with Switching Costs
9341	A. Demirgüç-Kunt and H. Huizinga	Barriers to Portfolio Investments in Emerging Stock Markets
9342	G.J. Almekinders	Theories on the Scope for Foreign Exchange Market Intervention
9343	E.R. van Dam and W.H. Haemers	Eigenvalues and the Diameter of Graphs
9344	H. Carlsson and S. Dasgupta	Noise-Proof Equilibria in Signaling Games
9345	F. van der Ploeg and A.L. Bovenberg	Environmental Policy, Public Goods and the Marginal Cost of Public Funds
9346	J.P.C. Blanc and R.D. van der Mei	The Power-series Algorithm Applied to Polling Systems with a Dormant Server
9347	J.P.C. Blanc	Performance Analysis and Optimization with the Power-series Algorithm
9348	R.M.W.J. Beetsma and F. van der Ploeg	Intramarginal Interventions, Bands and the Pattern of EMS Exchange Rate Distributions
9349	A. Simonovits	Intercohort Heterogeneity and Optimal Social Insurance Systems
9350	R.C. Douven and J.C. Engwerda	Is There Room for Convergence in the E.C.?
9351	F. Vella and M. Verbeek	Estimating and Interpreting Models with Endogenous Treatment Effects: The Relationship Between Competing Estimators of the Union Impact on Wages
9352	C. Meghir and G. Weber	Intertemporal Non-separability or Borrowing Restrictions? A Disaggregate Analysis Using the US CEX Panel

No.	Author(s)	Title
9353	V. Feltkamp	Alternative Axiomatic Characterizations of the Shapley and Banzhaf Values
9354	R.J. de Groof and M.A. van Tuijl	Aspects of Goods Market Integration. A Two-Country-Two -Sector Analysis
9355	Z. Yang	A Simplicial Algorithm for Computing Robust Stationary Points of a Continuous Function on the Unit Simplex
9356	E. van Damme and S. Hurkens	Commitment Robust Equilibria and Endogenous Timing
9357	W. Güth and B. Peleg	On Ring Formation In Auctions
9358	V. Bhaskar	Neutral Stability In Asymmetric Evolutionary Games
9359	F. Vella and M. Verbeek	Estimating and Testing Simultaneous Equation Panel Data Models with Censored Endogenous Variables
9360	W.B. van den Hout and J.P.C. Blanc	The Power-Series Algorithm Extended to the *BMAP/PH*/1 Queue
9361	R. Heuts and J. de Klein	An (s,q) Inventory Model with Stochastic and Interrelated Lead Times
9362	K.-E. Wärneryd	A Closer Look at Economic Psychology
9363	P.J.-J. Herings	On the Connectedness of the Set of Constrained Equilibria
9364	P.J.-J. Herings	A Note on "Macroeconomic Policy in a Two-Party System as a Repeated Game"
9365	F. van der Ploeg and A. L. Bovenberg	Direct Crowding Out, Optimal Taxation and Pollution Abatement
9366	M. Pradhan	Sector Participation in Labour Supply Models: Preferences or Rationing?
9367	H.G. Bloemen and A. Kapteyn	The Estimation of Utility Consistent Labor Supply Models by Means of Simulated Scores
9368	M.R. Baye, D. Kovenock and C.G. de Vries	The Solution to the Tullock Rent-Seeking Game When R > 2: Mixed-Strategy Equilibria and Mean Dissipation Rates
9369	T. van de Klundert and S. Smulders	The Welfare Consequences of Different Regimes of Oligopolistic Competition in a Growing Economy with Firm-Specific Knowledge
9370	G. van der Laan and D. Talman	Intersection Theorems on the Simplotope
9371	S. Muto	Alternating-Move Preplays and vN - M Stable Sets in Two Person Strategic Form Games

No.	Author(s)	Title
9372	S. Muto	Voters' Power in Indirect Voting Systems with Political Parties: the Square Root Effect
9373	S. Smulders and R. Gradus	Pollution Abatement and Long-term Growth
9374	C. Fernandez, J. Osiewalski and M.F.J. Steel	Marginal Equivalence in ν-Spherical Models
9375	E. van Damme	Evolutionary Game Theory
9376	P.M. Kort	Pollution Control and the Dynamics of the Firm: the Effects of Market Based Instruments on Optimal Firm Investments
9377	A. L. Bovenberg and F. van der Ploeg	Optimal Taxation, Public Goods and Environmental Policy with Involuntary Unemployment
9378	F. Thuijsman, B. Peleg, M. Amitai & A. Shmida	Automata, Matching and Foraging Behavior of Bees
9379	A. Lejour and H. Verbon	Capital Mobility and Social Insurance in an Integrated Market
9380	C. Fernandez, J. Osiewalski and M. Steel	The Continuous Multivariate Location-Scale Model Revisited: A Tale of Robustness
9381	F. de Jong	Specification, Solution and Estimation of a Discrete Time Target Zone Model of EMS Exchange Rates
9401	J.P.C. Kleijnen and R.Y. Rubinstein	Monte Carlo Sampling and Variance Reduction Techniques
9402	F.C. Drost and B.J.M. Werker	Closing the Garch Gap: Continuous Time Garch Modeling
9403	A. Kapteyn	The Measurement of Household Cost Functions: Revealed Preference Versus Subjective Measures
9404	H.G. Bloemen	Job Search, Search Intensity and Labour Market Transitions: An Empirical Exercise
9405	P.W.J. De Bijl	Moral Hazard and Noisy Information Disclosure
9406	A. De Waegenaere	Redistribution of Risk Through Incomplete Markets with Trading Constraints
9407	A. van den Nouweland, P. Borm, W. van Golstein Brouwers, R. Groot Bruinderink, and S. Tijs	A Game Theoretic Approach to Problems in Telecommunication

No.	Author(s)	Title
9408	A.L. Bovenberg and F. van der Ploeg	Consequences of Environmental Tax Reform for Involuntary Unemployment and Welfare
9409	P. Smit	Arnoldi Type Methods for Eigenvalue Calculation: Theory and Experiments
9410	J. Eichberger and D. Kelsey	Non-additive Beliefs and Game Theory
9411	N. Dagan, R. Serrano and O. Volij	A Non-cooperative View of Consistent Bankruptcy Rules
9412	H. Bester and E. Petrakis	Coupons and Oligopolistic Price Discrimination
9413	G. Koop, J. Osiewalski and M.F.J. Steel	Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function
9414	C. Kilby	World Bank-Borrower Relations and Project Supervision
9415	H. Bester	A Bargaining Model of Financial Intermediation
9416	J.J.G. Lemmen and S.C.W. Eijffinger	The Price Approach to Financial Integration: Decomposing European Money Market Interest Rate Differentials
9417	J. de la Horra and C. Fernandez	Sensitivity to Prior Independence via Farlie-Gumbel-Morgenstern Model
9418	D. Talman and Z. Yang	A Simplicial Algorithm for Computing Proper Nash Equilibria of Finite Games
9419	H.J. Bierens	Nonparametric Cointegration Tests
9420	G. van der Laan, D. Talman and Z. Yang	Intersection Theorems on Polytopes
9421	R. van den Brink and R.P. Gilles	Ranking the Nodes in Directed and Weighted Directed Graphs
9422	A. van Soest	Youth Minimum Wage Rates: The Dutch Experience
9423	N. Dagan and O. Volij	Bilateral Comparisons and Consistent Fair Division Rules in the Context of Bankruptcy Problems
9424	R. van den Brink and P. Borm	Digraph Competitions and Cooperative Games
9425	P.H.M. Ruys and R.P. Gilles	The Interdependence between Production and Allocation
9426	T. Callan and A. van Soest	Family Labour Supply and Taxes in Ireland

No.	Author(s)	Title
9427	R.M.W.J. Beetsma and F. van der Ploeg	Macroeconomic Stabilisation and Intervention Policy under an Exchange Rate Band
9428	J.P.C. Kleijnen and W. van Groenendaal	Two-stage versus Sequential Sample-size Determination in Regression Analysis of Simulation Experiments
9429	M. Pradhan and A. van Soest	Household Labour Supply in Urban Areas of a Developing Country
9430	P.J.J. Herings	Endogenously Determined Price Rigidities
9431	H.A. Keuzenkamp and J.R. Magnus	On Tests and Significance in Econometrics
9432	C. Dang, D. Talman and Z. Wang	A Homotopy Approach to the Computation of Economic Equilibria on the Unit Simplex
9433	R. van den Brink	An Axiomatization of the Disjunctive Permission Value for Games with a Permission Structure
9434	C. Veld	Warrant Pricing: A Review of Empirical Research
9435	V. Feltkamp, S. Tijs and S. Muto	Bird's Tree Allocations Revisited
9436	G.-J. Otten, P. Borm, B. Peleg and S. Tijs	The MC-value for Monotonic NTU-Games
9437	S. Hurkens	Learning by Forgetful Players: From Primitive Formations to Persistent Retracts
9438	J.-J. Herings, D. Talman, and Z. Yang	The Computation of a Continuum of Constrained Equilibria
9439	E. Schaling and D. Smyth	The Effects of Inflation on Growth and Fluctuations in Dynamic Macroeconomic Models
9440	J. Arin and V. Feltkamp	The Nucleolus and Kernel of Veto-rich Transferable Utility Games
9441	P.-J. Jost	On the Role of Commitment in a Class of Signalling Problems
9442	J. Bendor, D. Mookherjee, and D. Ray	Aspirations, Adaptive Learning and Cooperation in Repeated Games