Evolution towards efficient coordination in repeated games, preliminary version

Demichelis, Stefano

7 June 2012

Evolution towards efficient coordination in repeated games, preliminary version

Stefano Demichelis *

June 8, 2012

Abstract

We show that in long repeated games - or in infinitely repeated games with discount rate close to one- payoffs corresponding to evolutionary stable sets are asymptotically efficient, as intuition suggests. Actions played at the beginning of the game are used as messages that allow players to coordinate on Pareto optimal outcomes in the following stages. Strategies following some simple and intuitive "behavioral maxims" are shown to be able to drive out inefficient ones from a population. The result builds a bridge between the theory of repeated games and that of communication games that will be further investigated.

*To contact me please send an email at sdm.golem@gmail.com

1 Introduction

Consider the (doubly symmetric) game.

$$\begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 10 & 0 \\ 0 & 1 \end{pmatrix} \end{array} \tag{1.1}$$

and suppose you ask a friend, unacquainted with game theory, how two reasonable persons would play it. The likely reply would be that the obvious choice for both is A. In order to defend the usefulness of the solution concepts you have learnt, you may come out with a story to convince her that, in some cases, (B, B) could also be a conceivable outcome. To make the story short, suppose that she agrees that some "brutish power which, hidden, holds sway to common evil"[1], let's call it Momus, has convinced both players to play B, because "'This is what your coplayer will play because she believes that everybody does so. You cannot talk to her to change this convention. So if you play A you will get 0 instead of one"'. And, of course, if the the game where just a less symmetric one, such as,

$$\begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 10 & 0 \\ 8 & 7 \end{pmatrix} \end{array} \tag{1.2}$$

Momus's endeavors would be less toilsome. He would perfidiously remark that A may be an obvious choice but is certainly a risky one: because of the asymmetry of payoffs miscoordination will be more harmful to the A player than to the B player.

In either game, however, Momus's victory over common good (and maybe common sense, too) depends crucially on the lack of communication between players: if they could speak before the game they would be able to overcome their fears and agree on the "obvious" outcome [2].

But now your friend, deeply convinced that humans have a natural tendency to explore how poor cooperation may be improved, would point out that even without explicit communication the players could fool Momus and end up playing A. "Suppose", she would say " that they meet each other very frequently and that the game is played at every meeting - as is often

[1] "brutto poter che ascoso al commun danno impera"

[2] Game theory can be used to show this rigorously, as done in [16] and [9].

the case in real life - then, even if your Momus told them to play B all the time and even if communication is impossible, at some point they will switch to A, to their mutual advantage" "Why?" would you ask. She would reply "Because, if the game is repeated many times, the cost of miscoordination at one or even a few stages won't matter too much compared to the chance of getting 10 for the rest of the play. So it is very likely that one of them will deviate from Momus's prescription attempting to suggest a better way of playing. And, of course, the other will realize it: after all, they are both rational individuals, they know that the other one is rational too, if they follow the prescription they will get very little. So, as soon as there is a chance to move to a better regime, they will take advantage of it. I agree that this may be a little more difficult in game 1.2 compared to game 1.1: the outcome (A, B) is very bad for the one who plays A but may be acceptable to the B player. Still, in a frequently repeated game, I am confident that they would end up in (A, A). Actually it is not even necessary that they play the same game at each meeting; as long as in every stage the game has an obviously good outcome for both players, they will tend to agree on playing it as soon as possible. I would think that, using actions as messages, interpreting them as such and giving a coplayer the chance of expressing her doubts on suboptimal behaviors is an innate attitude of all of us".

The aim of this paper is to formalize and make rigorous this kind of argument. This not only in order to convince your friend that game theory has something to do with human behavior: the assumption that agents, whenever possible, coordinate on a Pareto efficient equilibria is a widespread one in economic theory[3]and it would be nice to have a theoretical foundation of it.

Let's begin with game 1.1 and repeat it N times, N large. What are the sensible payoffs that you would expect? The folk's theorem does not say too much. For instance it is easy to see that even a low average payoff like 1 can be supported by a Nash and even a perfect equilibrium [4]. Not only very bad outcomes are supported, but, if we look at the equilibria constructed to support them, we will be struck at how odd they are: you are supposed to blindly punish every kind of deviation, even those that benefit you and the punishment, in many games, may be more costly to you than to the one who

[3]This was, for instance, the assumption motivating [2] and much of the subsequent work on renegotiation proof equilibria.

[4]See [18] for references and extensive discussions on the topic

is deviating. (You have even to punish a coplayer for *not* punishing you!)[5].

Punishing somebody suggesting a mutually advantageous equilibrium does not seem to be a very successful attitude, especially from the evolutionary point of view. So you may wonder what evolutionary stability can say in these cases. At first sight the answer seems to be "as little as the folk theorem", in fact, it is easily seen that the strange (self) punishing equilibria above are neutrally evolutionary stable. The weakness is not in evolution, however, but in its modeling through neutral stability[6].

The problem here is that neutral stability overlooks a basic fact: you can punish deviations from the equilibrium path when they take a different action because you see them, but you cannot punish people who just *think* that they will not punish mutants[7]. More generally mutants deviating from the population on contrafactual events will undetectable and can not be selected away by any means. So there is no reason to believe that at some point they will be driven out from the population. Actually they could as well *invade* the population.

Neutral stability is blind to this phenomenon: indeed a population may be vulnerable to attacks by these "silent" mutants even if its strategy satisfies the formal requirement of NES. For this reason NES is usually very uninformative: in extensive form games, most components of equilibria contain NES, no matter how bad they are [8].

A better model of "evolution" or "learning" etc. is to consider how behavior changes under some adaptive process and identify the strategies, if any, on which it eventually settles, the technical term for them are asymptotically stable components of a dynamical system.

This is the approach of [7] where intrinsic conditions were found to insure that a component is asymptotically stable for at least one consistent payoff dynamic. As a consequence, "'intuitively bad"' components can be selected away as intrinsically evolutionary unstable. This result could, in principle,

---

[5]For experimental evidence that these are unnatural behaviors see [10] and the references therein.

[6]This was already noted, in an analogous case, by Fudenberg and Maskin in [12], where the difficulty was overcome assuming players are subject to involuntary mistakes during all the game

[7]Actually many would argue that, without explicit communication, it doesn't make sense to say how you would react if your coplayer did something you expect with probability zero. Not to say what you would do to a coplayer who just plans to do something different.

[8]A collection of paradoxical NES, with discussion of them will be available in [5]

be applied to our repeated game; however checking the conditions involves some elementary but nontrivial topological concepts and rather long proofs. To make the work accessible to as many game theorists as possible, this paper uses a more down to earth approach, based on the notion of ESSets, introduced by Thomas. If a set of Nash equilibria is a ESSet not only no mutant can do strictly better than the population but, if a mutant does as well as the population, it has to be already in the set. In this way the phenomenon of silent mutations described above is adequately dealt with. Formal definitions are in section 3.

Section 4 exposes the results of the paper. The main one is that ESSets consist of asymptotically efficient equilibria. Explicitly, when the game is repeated many times, say N, the average payoff of strategies in ESSets must converge to the one of the efficient outcomes in the stage game[9].

In the following paragraphs we sketch how the proof, given in section 5, proceeds. It illustrates what we think are some general principles of human behavior and our arguments give an evolutionary foundation for them. [10]. We show that, if a strategy is inefficient, the population adopting it can be invaded by mutants conforming to some simple behavioral maxims. Namely: use actions at the beginning of the play as messages and interpret them as such, do not punish those who communicate their willingness to deviate from inefficient equilibria and avoid "'babbling"'.

Actions, particularly at the beginning of the game, can be used to send messages; using and interpreting them in this way gives an evolutionary advantage. In our proof we exploit this in two ways: first an action different from those foreseen by the strategy can be used to signal that you are a mutant, second it can be used to try to generate asymmetric outcomes. So players can first agree that it is worthwhile to experiment better equilibria than the one they are on, than they can use the previous history as an anticoordination device, when anticoordination is needed. The first aspect comes into play in subsection 5.1: A strategy in an ESSet must have at least

[9]Our attention, in this paper, is focused on games such as Stag Hunting where the efficient outcome is a Pareto dominant equilibrium. For them the ideas at work can be expressed in a relatively simple form. Some results are however obtained and stated more generally and they will be used, in subsequent work on games of the Prisoner's Dilemma and Hawk-Dove type. In these cases the necessary conditions for evolutionary stability are the same but to find a concept that satisfies existence a more refined analysis is needed

[10]Some of these ideas where already *in nuce* in [16] and [9], in perspective they were already present, in a different form, in the seminal paper [12].

the same payoffs as one that is "nice to newcomers". Explicitly, a population will drift through silent mutations to one that has the same payoffs but, instead of punishing deviations from the equilibrium path, it rewards them by playing efficiently in the rest of the game. So a mutant could first deviate and then trigger optimal cooperation. If the population is playing something in an ESSet it must have a payoff at least as high as such a mutant.This approach is in its essence similar to the ones pioneered by Fudenberg and Maskin in [12].

But an inefficient strategy may try to defend itself by playing a babbling equilibrium: i.e. it would play all actions with nonzero probability long enough so as to make mutants unrecognizable or recognizable only when it is too late. This is where the anticoordination property of asymmetric histories becomes crucial. In section 5.2, "do not babble", we show that strategy in an ESSset can drift to less noisy ones. This is less easy as at a superficial sight would seem, as the examples at the beginning of the subsection show [11].

Finally what should do if the strategy is "babbling" and the previous history is symmetric? The answer is "nothing": in section 5.3, "be patient", we show that such a case cannot repeat itself with too high probability, so its contribution to the expected payoffs of strategies can be neglected.

In the end it is shown that all payoffs in an ESSet must be at least as high as a strategy that devotes a certain number of stages $c(N)$ to communication only, getting the lowest possible payoff in the game, and then gets the Pareto efficient payoff. This is our main result.

It is interesting to study $c(N)$ more in detail. Not only does the ratio $c(N)/N$ go to zero, implying our asymptotic efficiency, but its order of magnitude is rather small, i.e. $N^{1/2}$. In an important special case it becomes even smaller, just one, this is when the game is doubly symmetric, i.e. the players have the same payoffs. There is a natural explanation of this fact: if the players' interests are totally aligned, it is necessary only one round of the game to convey the message "this equilibrium is suboptimal, we can do better from now on, let's do it" and it is immediately understood. An example shows that the condition of double symmetry is crucial, in games

[11] The problem is that, even if one begins with a game such a Stag Hunting, with Pareto dominant evolutionary stable equilibria, in the repeated game some strategies can generate strategic situations of the Hawk Dove game, with only one evolutionary stable symmetric equilibrium, as explained in the example at the beginning of section 5.2, see also 5.3 in the appendix

where different payoffs generate a conflict between Pareto dominance and risk dominance more rounds are needed...players do not trust each other at first sight [12].

In this discussion we dealt with finitely repeated games. It is intuitive that actual human beings should play infinitely repeated games with discount close to one in a similar way to long finite games, even if many solutions concept are not robust with respect to this analogy. The results of this paper have this property and the mechanisms underlying the proofs are essentially the same as shown in section 6.

Another feature of the function $c(N)$ is that it is a *universal* function, it depends only on N so in *all* games no more than $c(N)$ stages are needed to agree on playing efficiently, no matter how bad the priors on the other player's behavior are. Moreover, this this allows to make our result stronger, as done in section 7. We allow the stage game to change from one step to the other, provided that its payoffs stay bounded and there is always a unique Pareto dominant equilibrium. The same results as for fixed stage games holds.

This fact is, in my opinion, important because it answer a potentially fatal objection to the evolutionary approach: "Long repetitions of the same game happen seldom in life so, how can the evolutionary force be strong enough to generate an appreciable speed of evolution?". The answer is that, even if the same game with the same partner is seldom repeated, we are often matched to partners, friends, colleagues with whom we play different games at different times, and often they are all cooperation games, at least approximately. According to our results, in these cases too evolution selects strategies that do at least as well as those respecting some "behavioral maxims", like the ones used in the proof. "If you are locked in an inefficient social rule and your partner, whose interests are aligned with yours, deviates from it, be kind!: she may suggest the "better way" you were yearning to", "Avoid confusing actions, you must be able to recognize innovators". They may be philogenetically evolved rule of thumbs to optimize our interactions and this can explain why we find them so "natural" while equilibria violating them look so "artificial" or "strange".

The expression "behavioral maxims" is an intentional reference to Grice's conversational maxims [13]. In the same way as optimal cooperation in a

[12] See[1]for a discussion of the problem of "'trusting"' in the one stage game

[13] See [14] and, for a textbook exposition, [17]

conversation requires the use of certain conversational maxims, evolutionary successful players in repeated games must respect similar behavioral maxims, they are the rules for effective implicit communication. [14]

An effort has be done to keep the exposition to a level accessible to most game theorists. The proofs are often not straightforward and require some willingness to carefully concentrate on details but the prerequisites do not go beyond the definition of limit. The ones conceptually more relevant are in the paper, the more technical ones in the Appendix. In order to keep the size of this contribution within reasonable limits, calculus exercises and straightforward checks are left to the reader.

2 The class of games

In this section we set the notation and we define the class of games we will study: symmetric two players games that will be repeated a finite number of times, denoted by N.

2.1 Stage Game

We begin with the stage game. Let $G = (\mathbf{A}, u)$ be a symmetric two-players game, where $\mathbf{A}$ is the finite set of pure strategies available to each player, and $u : \mathbf{A}^2 \to \mathbf{R}$ is the payoff function: for $a, b \in \mathbf{A}$, $u(a, b)$ is the payoff to a player who uses pure strategy a against pure strategy b. Let $\Delta(\mathbf{A})$ denote the set of mixed strategies, that is, the unit simplex spanned by $\mathbf{A}$, then u extends to a function $u : [\Delta(\mathbf{A})]^2 \to \mathbf{R}$ in the usual way.

Note the difference with respect to an (asymmetric) two-players game $(\mathbf{A_1}, \mathbf{A_2}, u_1, u_2)$. In this case if player 1 plays a and player 2 plays b, the payoff to player 1 is $u_1(a, b)$ and the payoff to player 2 is $u_2(a, b)$ (not $u_2(b, a)$!). So a symmetric game is a two-players game such that $u_1(x, y) = u_2(y, x)$ and this quantity is denoted by $u(x, y)$. We also remind the reader that a doubly symmetric game is a symmetric game in which the two players get the same payoff, i.e. $u(x, y) = u(y, x)$. From now on strategies in the symmetric game will be called "Actions".

[14] This is more than an analogy: in [6] it will be shown how at least a fragment of Grice's maxims can be translated in necessary and sufficient conditions for efficient communication in repeated games

Definition 2.1. *Given a symmetric stage game G, let* $\max_{x\in A} u(x,x) = P$, $\max_{x,y\in A} \frac{u(x,y)+u(y,x)}{2} = R$ *and* $\min_{(x,y))\in A^2} u(x,y) = Q$, *$P$ will be called the* symmetric efficient payoff, *R will be called the* efficient payoff *and Q will be called the* worst payoff. *The difference $D = P - Q$ will be called the* worst loss. *We say that an action a such that $a \in \operatorname{Argmax}_{x\in A} u(x,x)$ is an* optimal action. *An action pair (b,c) such that $(b,c) \in \operatorname{Argmax}_{x,y\in A} [u(x,y)+u(y,x)]$ is an* optimal action pair

There are two generic cases:

1. Either $P = R$, so there is a single action, the optimal one, realizing the efficient payoff as in Stag Hunting,
2. or $P < R$, so the efficient payoff obtains in an antisymmetric outcome, as in the Hawk-Dove game.

An important class of generic symmetric games is the one in which the efficient payoff is not only symmetric but also a Nash equilibrium, it is the object of the next definition.

Definition 2.2. *We say that a symmetric game (G,u) is a* Paretian Game *if the outcome of an optimal action (a,a)* strictly *Pareto dominates all the asymmetric ones, i.e.*

$$u(a,a) = \max_{(x,x)\in\mathbf{A}} u(x,x) > \max_{(x,y)\in\mathbf{A}} u(x,y)$$

Note that we do not require the optimal action to be unique, each optimal action gives a symmetric strict Nash equilibrium (a,a). Examples of Paretian Games are coordination games, e.g. stag hunting, but also more general ones such as the one given below.

$$\begin{array}{c c} & \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \end{array} & \begin{pmatrix} 10 & 0 & 3 & 1 \\ 9 & 10 & 0 & 8 \\ 9 & 8 & 2 & 9 \\ 9 & 8 & 7 & 4 \end{pmatrix} \end{array}$$

Exercise 2.1. *Find all Nash equilibria in the game, discuss their equilibrium refinement and evolutionary properties. What would be a "'reasonable"' way of playing it, according to you?*

2.2 Finitely repeated games

Now we think of repeating N times the stage game G , we will denote the repeated game by ${}^{N}G$. The notation developed below is adapted from the one in [18], some changes are principally due to the fact that time begins at one and is finite.

Time: Let the set of time periods, t, be the $\{1....N\}$.

Perfect monitoring: Assume *perfect monitoring* in the sense that all actions in earlier periods are observed before the current period's actions are taken (simultaneously).

Histories: Let the set of *histories* in ${}^{N}G$ be

$$H = \{\aleph\} \cup \bigcup_{t=1}^{t=N-1} H_t \tag{2.1}$$

it is the union of a one element set, the zero history, $h_0 = \aleph$, and the $H_t = (\mathbf{A^2})^t$ for each $t \geq 1$. For each period t, H_t is the set of all possible action profiles that might have been taken up to t included: $[(a_1, b_1), (a_2, b_2), ..., (a_t, b_t)]$. A history is thus a finite string of action pairs, across all earlier periods. Each such history uniquely defines a subgame, and each subgame uniquely defines a history. Given a history at time t , $h_t = [(a_1, b_1), (a_2, b_2), \ldots, (a_t, b_t)]$, we define its mirror history as $\bar{h}_t = [(b_1, a_1), (b_2, a_2), \ldots, (b_t, a_t)]$, this is the history obtained by reversing the role of the two players. If you have played the as and your opponent has played the bs, you will see the h_t above and your opponent will see $\bar{h}_t$. If $h_t = \bar{h}_t$ we will say that h_t is symmetric. We will sometime have to use partial histories defined on segments of time, e.g. $k_{t,s} = [(a_t, b_t), \ldots, (a_s, b_s)]$. Given a history $h_t = [(a_1, b_1), (a_2, b_2), \ldots, (a_t, b_t)]$ and a couple of actions (a, b) we define the composed history $h_t \circ (a, b) \in H_{t+1}$ as $h_t \circ (a, b) = [(a_1, b_1), (a_2, b_2), \ldots, (a_t, b_t), (a, b)]$, in the same way given another history $k_{t+1,s} = [(a_{t+1}, b_{t+1}), \ldots, (a_s, b_s)]$ we define $h_t \circ k_{t+1,s} = [(a_1, b_1), \ldots, (a_t, b_t), (a_{t+1}, b_{t+1}), \ldots, (a_s, b_s)]$. If $g_s = h_t \circ k_{t+1,s}$ for some $k_{t+1,s}$,we will say that h_t is an "ancestor" of g_s and denote this relation by $h_t \triangleright g_s$.

Strategies: A *behavior strategy* σ for a player is a mapping from histories to randomized actions in G:

$$\sigma : H \to \Delta(\mathbf{A}). \tag{2.2}$$

So, if, at time t, you see history h_t, strategy σ is telling you to play the randomized action $\sigma(h_t)$ and if your opponent is using strategy τ, she will be playing $\tau(\bar{h}_t)$ against you. Note that, if h_t is an asymmetric history, we will have in general $\tau(\bar{h}_t) \neq \sigma(h_t)$, even if τ coincides with σ. Given an action $a \in A$ we will denote with $\sigma(h_t)(a)$ the probability that a is played at t, conditional on h_t. The space of behavior strategies will be denoted by ${}^N\mathcal{B} = \prod_{h \in H} \Delta(\mathbf{A})$, a product of simplexes. Strategies prescribing pure actions at each history will be called pure.

Plays: A play $r_t = [(a_1, b_1), \ldots, (a_t, b_t)]$ is an element of $(\mathbf{A}^2)^t$, it is what has been played up to period t. For general information structures, it is advisable to distinguish plays and histories, as done in [18]. When perfect monitoring is assumed, as in this paper, they coincide.

Probabilities: Each behavior-strategy profile (σ, τ) recursively defines a probability distribution over the plays, as follows. An application of σ and τ to $\aleph$ defines a probability distribution over the set of action pairs in period 1. For each such realization, $h_1 = (a_1, b_1)$, an application of σ and τ to h_1 defines a probability distribution over the set of actions in period 2, etc. The measure projects on partial plays. Probabilities for histories are defined in the same way. Even when the probability of a history is zero, probabilities can be conditioned on it without ambiguity. The probability that play r_t will be played by the strategy profile (σ, τ), will be denoted by $p^{(\sigma,\tau)}(r_t)$. We will also need the probability that the partial play $r_{t+1,N} = [(a_{t+1}, b_{t+1}), \ldots, (a_N, b_N)]$ is played, conditional on on history h_t being realized, it is written as $p_{h_t}^{(\sigma,\tau)}(r_{t+1,N})$, note that it is well defined even if h_t has zero probability. When the two strategies coincide we will write $p^\sigma(h_t)$ for $p^{(\sigma,\sigma)}(h_t)$. Note that, by exchanging the two players, $p^{(\sigma,\tau)}(h_t) = p^{(\tau,\sigma)}(\bar{h}_t)$ and so $p^\sigma(h_t) = p^\sigma(\bar{h}_t)$.

Payoffs: Given a play $r_t = [(a_1, b_1), ..., (a_t, b_t)]$ its payoff is

$$U(r_t) = \sum_{i=1}^{t} u(a_i, b_i) \tag{2.3}$$

it is what a player playing the a's earns against the one playing the b's from period 1 to period t inclusive. In a similar way we define $U(h_t)$

for a history. The payoff of strategy σ against strategy τ, denoted by ${}^{N}U(\sigma, \tau)$, will be :

$$ {}^{N}U(\sigma, \tau) = \sum_{r_t} p^{(\sigma,\tau)}(r_N) U(r_t) \tag{2.4} $$

We will also use the *conditional payoff*

$$ {}^{N}U_{h_t}(\sigma, \tau) = \sum_{r_{t+1,N}} p_{h_t}^{(\sigma,\tau)}(r_{t+1,N}) U(r_{t+1,N}) \tag{2.5} $$

where

$$ U(r_{t+1,N}) = \sum_{i=t+1}^{N} u(a_i, b_i) \tag{2.6} $$

it is the payoff that σ expects against τ in the subgame defined by history h_t. We will simply write ${}^{N}U(\sigma)$ and ${}^{N}U_{h_t}(\sigma)$ for ${}^{N}U(\sigma, \sigma)$ and ${}^{N}U_{h_t}(\sigma, \sigma)$ respectively.

In many of the proofs and in some examples we will use a family of auxiliary games, called forward games, described in Appendix A, they will be our main technical tool. The reader who wants to follow the proof in detail should have now at least quick look at it.

3 Evolutionary stability concepts

Let now fix G, the stage game and ${}^{N}G$ the N repeated game. Remember that ${}^{N}\mathcal{B} = \prod_{h \in H} \Delta(A)$ is the space of behavior strategies for ${}^{N}G$.

The original definition of Evolutionary Stable Strategy was given in [19]:

Definition 3.1. *A strategy σ is* evolutionary stable *in a symmetric game (G, A, U) if:*

$$ \forall\ \tau U(\tau, \sigma) \leq U(\sigma, \sigma) \tag{3.1a} $$

$$ U(\tau, \sigma) = U(\sigma, \sigma) \Rightarrow U(\tau, \tau) < U(\sigma, \tau) \tag{3.1b} $$

this definition is very satisfactory for generic games but is, in general, not satified if the game is not generic, as is the case of our repeated games. in order to prove that something is something the strict inequality in 3.1b is often weakened to a weak inequality, giving the so called neutrally evolutionary strategies (NES). Being a neutrally evolutionary strategy is rather unifromative, as argued in the introduction: a collection of "strange" equilibria that are nevertheless NES is in [5] and available from the author.

A better approach to evolutionary stability is given by ESSets, as introduced by Thomas in [20], its definition, adapted to behavior strategies, is:

Definition 3.2. *A non-empty and closed set $X \subset {}^N\mathcal{B}$ is an* evolutionarily stable set *(an ESSet) for ${}^NG$ if for each $\sigma \in X$ there exists some $\delta > 0$ such that $u(\sigma, \sigma') \geq u(\sigma', \sigma')$ for all σ' in the best reply to σ within distance δ from it, with strict inequality if $\sigma' \notin X$.*

Intuitively when a mutant plays the best reply to the population and, when in the environment it creates gets the same payoffs as population members do, this mutant must be in the set. Next another definition:

Definition 3.3. *We say that σ' is an* elementary mutation *of σ at h_t if σ' differ from σ only on the history h_t, or its mirror $\bar{h}_t$, i.e $\sigma(k_s) = \sigma'(k_s)$ if $k_s \neq h_t, \bar{h}_t$.*

Elementary mutations do not require mutants to coordinate their changes across periods and so are, in a sense, the simplest and likeliest to occur.

Proposition 3.1. *Let $\sigma \in X$, X ESSet and let σ' be an elementary mutation of σ at h_t, then:*

$$ {}^NU(\sigma', \sigma) \leq {}^NU(\sigma, \sigma) \tag{3.2a}$$

$$ \textit{if } {}^NU(\sigma', \sigma) = {}^NU(\sigma, \sigma) \textit{ then } {}^NU(\sigma', \sigma') \leq {}^NU(\sigma, \sigma') \tag{3.2b}$$

$$ \textit{if } {}^NU(\sigma', \sigma) = {}^NU(\sigma, \sigma) \textit{ and } {}^NU(\sigma', \sigma') = {}^NU(\sigma, \sigma') \textit{ then } \sigma' \in X \tag{3.2c}$$

The proof, immediate and left to the reader, is the same as the well known one for normal form games. [15]

[15] Note that for more elaborate mutations it does not apply, due to the multilinearity of the payoff function.

This is the only property of ESS that we will use, so we will prove something stronger than our statement: our lower bound on payoffs will hold for also for sets just immune to attacks on one point of the strategy at a time.

We make this fact explicit in the following

Definition 3.4. *We say that a closed set X of behavior strategies is an ESSp if for $\sigma \in X$, and σ' an elementary mutation of σ at history at h_t, we have:*

$$^N U(\sigma', \sigma) \leq {}^N U(\sigma, \sigma) \tag{3.3a}$$

$$\textit{if } {}^N U(\sigma', \sigma) = {}^N U(\sigma, \sigma) \textit{ then } {}^N U(\sigma', \sigma') \leq {}^N U(\sigma, \sigma') \tag{3.3b}$$

$$\textit{if } {}^N U(\sigma', \sigma') = {}^N U(\sigma, \sigma') \textit{ and } {}^N U(\sigma', \sigma) = {}^N U(\sigma, \sigma) \textit{ then } \sigma' \in X \tag{3.3c}$$

So, by proposition 3.1, every ESSet set is an ESSp. In this paper assumptions will be always the minimal ones, i.e. that the set is an ESSp, and conclusions will be the strongest ones, i.e. that the set is an ESSet. Note that ESSets consist of Nash equilibria and ESSp sets induce Nash equilibria in every forward game, this is exercise A.4 in the appendix [16].

It is not difficult to see that a singleton set $X = \{\sigma\}$ is an ESSet if and only if the strategy σ is an evolutionary stable one.

An important application of proposition 3.1 is to mutations that do change the behavior of σ only on zero probability histories. The precise definition is

Definition 3.5. *Let σ be a strategy in ${}^N G$, we say that σ' is a* silent mutation *of σ if*

$$p^\sigma(h_t) \neq 0 \Rightarrow \sigma'(h_t) = \sigma(h_t) \tag{3.4}$$

The important property of silent mutation is stated in the next proposition

[16] There is a lot of leeway in the choice of definitions of evolutionary stability in extensive form games. One can argue about populations and mutants coordinating their actions across periods or not, express the conditions in mixed or behavioral strategies, discuss their normal and agent normal forms etc. This would be an entertaining exercise, in the spirit of [8]. As it goes outside the scope of this paper it will not be done here, that is why we state results in their strongest form.

Proposition 3.2. *If σ' is a silent mutation of σ then*

$$\forall h_t,\ p^{\sigma}(h_t) = p^{\sigma'}(h_t) = p^{(\sigma,\sigma')}(h_t) = p^{(\sigma',\sigma)}(h_t) \tag{3.5a}$$

$$^{N}U(\sigma,\sigma) = {}^{N}U(\sigma',\sigma) = {}^{N}U(\sigma,\sigma') = {}^{N}U(\sigma',\sigma') \tag{3.5b}$$

$$\sigma \in X,\ X\ ESS \Rightarrow \sigma' \in X \tag{3.5c}$$

$$\sigma \in X,\ X\ ESSp \Rightarrow \sigma' \in X \tag{3.5d}$$

The proof of equations (3.5a) and (3.5b) is obvious from the definition of silent mutation. Implication (3.5c) follows from (3.5d) because every ESSet is a ESSp. To prove (3.5d) we use induction: choose some ordering of the zero probability histories and mutate σ step by step so as to obtain a sequence $\sigma_0 = \sigma, \sigma_1, ..., \sigma_n = \sigma'$ of silent mutations such that each σ_{i+1} is an elementary mutation of σ_i. The proof works because each elementary silent mutation does not change the probability of histories. Then apply proposition 3.1 at each step using equations (3.5a) and (3.5b).

Definition 3.6. *We say that σ' is a* submutation *of σ if*

$$^{N}U(\sigma',\sigma) = {}^{N}U(\sigma',\sigma) \tag{3.6a}$$

$$^{N}U(\sigma',\sigma') = {}^{N}U(\sigma',\sigma') \tag{3.6b}$$

$$^{N}U(\sigma',\sigma') \leq {}^{N}U(\sigma,\sigma) \tag{3.6c}$$

Conditions (3.6a) and (3.6b) together imply that if $\sigma \in X$, an ESSp set, then $\sigma' \in X$, too.

A submutation is a dangerous and silly mutant: it can enter the population by making both the population and itself ... worse. Still they will prove useful to us because, if we can find a lower bound for the payoff of the submutation, this will hold for the original strategy too.

Definition 3.7. *Given a game G and the corresponding repeated game ^{N}G, we say that π is an* ESS payoff of *^{N}G if there exist an ESSet X for ^{N}G and a $\sigma \in X$ such that $^{N}U(\sigma) = \pi$.*

We say that $\bar{\pi}$ is an average ESS payoff *for ^{N}G if there is an ESSet X for ^{N}G and a $\sigma \in X$ such that $^{N}U(\sigma)/N = \bar{\pi}$.*

Corresponding definitions hold for ESSp payoffs.

4 Results for finitely repeated games

We can now state our results. The main technical one is proposition 4.1 that shows that the average ESS payoffs for ${}^N G$ are asymptotically efficient, i.e. they cannot be smaller than the symmetric efficient payoff in the limit of N infinity. Actually we prove much more and we find a universal function $c(N)$ measuring the convergence rate for all games.

Moreover, if σ is a strategy in an ESSp set that uses only pure actions the bound on the convergence rate is dramatically improved, it is just 1, and the proof is particularly simple. We state the result separately in proposition 4.2.

A more intriguing drama betides when the game is doubly symmetric, there too $c(N)$ is one, as discussed in subsection 4.1.

Proposition 4.1 is then applied in theorem 1 to characterize ESS payoffs in paretian games. Extensions and further applications will be discussed in the conclusions.

Proposition 4.1. *There is a universal function $c(N)$ with $\lim_{N\to\infty} \frac{c(N)}{N^{1/2+\varepsilon}} = 0$ for all $\varepsilon > 0$, such that:*

If G is a stage game with efficient symmetric payoff P and with worst loss D and if $\Pi(N) = \inf \left\{ {}^N U(\sigma) | \sigma \in X,\ X \text{ } ESSet \text{ } for \text{ } {}^N G \right\}$ then:

$$\Pi_N \geq P \cdot N - c(N) \cdot D \tag{4.1}$$

in particular, the average ESS payoff will be approximately P or larger when N goes to infinity. More precisely, if X_N is a sequence of ESS for ${}^N G$ and,for every N, $\sigma_N \in X_N$, we have:

$$\liminf_{N\to\infty} \frac{{}^N U(\sigma_N)}{N} \geq P \tag{4.2}$$

In case the strategy is a pure one we can say even more, and the proof is much easier: as said $c(N)$ turns out to be one:

Proposition 4.2. *If X is an ESS for G_N and if $\sigma \in X$ uses only pure actions :*

$${}^N U(\sigma) \geq N \cdot P - D \tag{4.3}$$

The function $c(N)$ has an interesting interpretation: ESS payoffs are *as if*, out of N rounds, $c(N)$ of them where used exclusively to convey messages between players then the optimal action is used.

In this light it is clear that, if the strategy is a pure one, just one round of communication should be needed:if the equilibrium is inefficient, the mutant makes herself recognizable, players accept her implicit proposal and then play efficiently.

If strategies are mixed something more interesting happens: the order of magnitude of $c(N)$, approximately $N^{1/2}$, is the same as the order of magnitude of the average standard deviation of statistics on actions. It is as if the players where trying to understand if the deviations of their partner are due to random fluctuations or are intentional attempts to convey a message. This is, for the moment, only a suggestive interpretation: the *a priori* probability of a mutant in our model is zero so the mutant can be recognized only after that other mutations have made the strategy drift to a non completely mixed one and the mutant plays an action outside the support. We plan to investigate the question in further work, where a continuous stream of non zero probability mutants will be assumed.

Theorem 1. *Let G be a Paretian game with optimal action a, optimal payoff P and worse loss D and let its N repetition be ${}^N G$. Then*

1. *The (possibly disconnected) set $\{\sigma \in {}^N\mathcal{B} | {}^N U(\sigma) = N \cdot P\}$ consisting of strategies playing always an optimal action against themselves is an ESS.*
2. *If $\sigma \in X$, X an ESSp, then ${}^N U(\sigma) \geq N \cdot P - c(N) \cdot D$*

Here $c(N)$ is the same function as in proposition 4.1.

Proof: The existence part is easy, if the game is paretian no action or couple of actions can get a higher payoff than P and P obtains only when an optimal action is played against itself. So, if $\sigma \in X$ and $\tau \notin X$ one has the strict inequality ${}^N U(\tau, \sigma) < {}^N U(\sigma, \sigma)$ and the result follows. Part two is proposition 4.1 ■

For existence of ESSets the strictness of the optimal action required in definition 2.2is necessary as the following example, a borderline case between Prisoner's Dilemma and Stag Hunting, shows:

$$\begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 10 & 5 \\ 10 & 7 \end{pmatrix} \end{array} \tag{4.4}$$

The reader can, as usual, prove it as an exercise. The theorem says that the repeated game has at least an ESS and that all of them are approximately efficient. Note that, even if the optimal action is unique, there are in general other ESSets besides playing strictly optimally. For instance if G is generic, a is the unique optimal action and (b, b), $b \neq a$ is another strict Nash equilibrium, the set $X_i = \{\sigma | \sigma(h_t) = a \text{ if } t \neq i \text{ and } p^\sigma(h_t) > 0 \text{ while } \sigma(h_i) = b\}$ of strategies playing a on the equilibrium path at all times but at i is an ESS.

When the efficient payoff R is larger than P and so is given by a couple of different actions, the results can be improved. In fact asymptotic efficiency would consist first in trying to produce an asymmetric history, so that players can be assigned roles, and then playing the asymmetric efficient equilibrium. This is the object of next proposition.

Proposition 4.3. *There is a universal function $d(N)$ with $\lim_{N\to\infty} \frac{d(N)}{N^{1/2+\varepsilon}} = 0$ for all $\varepsilon > 0$, such that: if G is a stage game with efficient payoff R and with worst loss D and if π_N is an ESS payoff of ${}^N G$ then:*

$$\pi_N \geq R \cdot N - d(N) \cdot D \tag{4.5}$$

The proof is an adaptation of the one of proposition 4.3. It will not be given in the paper because the proposition, in itself, is of little use without an existence result, as discussed in section 8

4.1 Doubly symmetric games and an instructive counterexample

For doubly symmetric games, i.e. those in which each outcome gives the same payoff to the two players, intuition suggests that the evolution towards optimal cooperation should be quicker, given that interests are totally aligned. In fact the measure of communication inefficiency $c(N)$ is 1.

Theorem 2. *If the stage game G is a doubly symmetric one and if $\sigma \in X$, X an ESSp for ${}^N G$, then:*

$${}^N U(\sigma) \geq N \cdot P - D \tag{4.6}$$

For the proof, whose details are given in the Appendix, it is crucial that the game is doubly symmetric. The following counterexample shows that it cannot extended to just symmetric games and illustrates how games with "risky" strategies require longer repetitions.

Let the stage game be

$$\begin{array}{c} \\ a \\ b \\ c \end{array}\begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 10 & 1 & 1 \\ 9 & 0 & 3 \\ 9 & 3 & 0 \end{pmatrix} \end{array} \tag{4.7}$$

and repeat it two times.
Consider the strategy:

$$\sigma(h_t)) = \begin{cases} 1/3a + 1/3b + 1/3c & \text{if } h_t = h_0 = \aleph \\ 1/2b + 1/2c & \text{if } h_t = (a,a), (b,b) \text{ or } (c,c) \\ c & \text{if } h_t = (a,b), (a,c) \text{ or } (b,c) \\ b & \text{if } h_t = (b,a), (c,a) \text{ or } (c,b) \end{cases} \tag{4.8}$$

It is an isolated Nash equilibrium, and it is relatively inefficient and in particular ${}^2U(\sigma) = 6.5 < 2 \cdot P - D = 2 \cdot 10 - 10 = 10$. Nevertheless as a set it is an ESSet, actually it is even a strictly evolutionary stable strategy. The formal proof of these facts is in the appendix, here we just sketch the key point, to illustrate the role of asymmetry in making it stable albeit relatively inefficient.

Let's consider how a mutant, τ, could drive the population out by deviating in the first stage only (in the appendix it will be seen that changes in stage two make the arguments even stronger). To the payoffs in stage one we add the payoffs that would be obtained after the corresponding history in stage two assuming the original strategy is respected (this is the forward game, see appendix A for the formal definition.)

$$\begin{array}{c} \\ a \\ b \\ c \end{array}\begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 11.5 & 4 & 4 \\ 12 & 1.5 & 6 \\ 12 & 6 & 1.5 \end{pmatrix} \end{array} \tag{4.9}$$

Suppose that, a mutant τ is trying to improve on σ. It would be tempting to raise the weight of playing a in the first stage. Indeed if $\tau(\aleph) = a$ and τ coincides with σ in the stage two, we have ${}^2U(\tau,\tau) = 11.5 > 6.5 = {}^2U(\sigma,\sigma)$. Unfortunately ${}^2U(\sigma,\tau) = 11 + 5/6$ is still greater, and so τ will not be able to enter the population.

The conceptual reason for this is that, due to the lack of double symmetry, a is a good but risky strategy: if you play a and the opponent doesn't, she gets a hight payoff while you loose a lot, and since the population is playing all actions with the same probability this will be likely to happen.

So a mutant naively deviating to a will not be able to drive the population out, the benefit it gives to the population are higher than the benefit it gives to itself. In fact the stage game contains a stag hunting game and the forward game even a prisoner's dilemma.

5 Proofs

In this section we prove of propositions 4.1 and 4.2.

We start from an arbitrary strategy σ in an ESSp, X. The set X will be kept fixed through all the proof. We will first move σ in X via silent mutations and submutations until we we will reach a new strategy σ', particularly easy to study. Than we will show that the payoff of this one is higher than $N \cdot P - c(N)D$. Since silent mutations do not change ${}^{N}U(\sigma)$ and submutation may only decrease it, any σ in X satisfies the lower bound, too.

The argument consists of three parts, each of them related to a different aspect of evolution.

In subsections 5.1 we show that the strategy can evolve within X into an *open minded* one, i.e. one that upon recognizing a mutant kindly plays the optimal action. The process does not change its payoff. This is already enough to show that, in X, strategies with non-full support at the early stages of the game are efficient.

In subsection 5.2 we deal with strategies with full support. We concentrate on what happens after an asymmetric history. It is possible, without leaving X of course, to exploit the history as an (anti)coordination device and to assign roles to the players, so that they employ asymmetric actions against each other. This reminds, in its simpler form, the passage from the symmetric mixed strategy to the asymmetric pure strategies in the Hawk-Dove game. There is, however, an important additional fact: our goal is not to increase the payoffs but to reduce the support of the strategy and then to apply the results of subsection 5.1. In fact, we want to avoid precisely that when a mutant enters the population, both the population and the mutant payoffs increase by the same amount. This is the object of a detailed and technical analysis in the appendix, where we show that it is possible to decrease the support using submutations only.

In subsection 5.3 we deal with the remaining case: strategies that have full support after symmetric histories. Since they have full support, there is a nonzero probability that the players play different strategies and so generate an asymmetric outcome to which the reasonings in subsection 5.2 can be applied. We estimate this probability in terms of the payoffs and we show that low payoffs imply high probability of generating symmetric outcomes.

A final argument, given in subsection 5.4, uses the optional sampling theorem, the conditional payoffs are recognized as a supermartingale that, stopped at an appropriately chosen time, satisfies the lower bound we want. An elementary proof avoiding martingale theory is also sketched.

At the beginning of the first three subsections we discuss informally some examples in order to introduce the concepts that will be used.

5.1 Be kind to newcomers

In this subsection the focus is on pure actions. Let's see first in an example how they lead to efficient play.

Consider the stage game

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 10 & 0 \\ 0 & 7 \end{pmatrix} \end{array} \tag{5.1}$$

and suppose we repeat it N times with N at least 4. It should be obvious that there are ESSets for the game, e.g. playing C all the time and getting $10N$, the best possible payoff.

We claim that if X is an ESS set and σ a *pure* strategy in it, the payoff of σ cannot depart too much from efficiency, in fact, σ must earn at least $10(N-1)$ against itself. To see this, let σ' be the strategy coinciding with σ except that it is "'open minded"', i.e. if it sees something outside the equilibrium path -and so realizes that the coplayer is a mutant- it will play C against it. It is clear that σ' is a silent mutation of σ and so, by proposition 3.2, we have $\sigma' \in X$, earning the same payoff as σ.

Let now τ be a mutant that in the first stage plays an action different from the one played by σ'; this is possible because we assumed that σ, and so σ' plays pure actions. In all other stages τ plays C unconditionally. Let's see what happens to τ when it meets σ'. At $t = 1$, it earns 0, suffering the worst possible loss in G. Still deviation was an insightful choice: now σ' has recognized it and, being open minded, is nice to it (and to itself...) by playing

C; so, at $2, 3, \ldots, N$ it earns again 10. Altogether this makes $(N-1)10$. So, if σ' has to be a best reply to itself it must earn at least as much, and the same holds for σ.

This simple mechanism is used by propositions 5.1 etc. below to prove the general case and something more. It is enough, indeed, that one of the $\sigma(h_t)$ has not full support in some early stage of the game for the conditional payoff U_{h_t} to be high, the sooner this happens the higher the conditional payoff will be.

Note that our result does not say anything about how the high payoff is reached. In the up to 99 times repeated stage game:

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 10 & 0 \\ 0 & 9.9 \end{pmatrix} \end{array} \tag{5.2}$$

the strategy playing, (D, D) all the time is in an ESSet. Of course, if G besides being paretian is also generic, the optimal action a will be unique and so, when N goes to infinity, the only way to achieve high payoffs will be to play a often enough.

We now begin with the formal proof.

Let G be a symmetric game and with a, P and Q and D as in definition 2.1. First note that we can subtract Q from all the payoffs and divide the result by D. This does not change the structure of Nash equilibria or ESSets, of course, and rescales all payoffs in sight. So, without loss of generality, we will assume in the proofs of this section that $P = 1$, $Q = 0$ and $D = 1$ [17]. Explicitly:

Assumption 5.1. *1. G is a symmetric game, with $P = 1$, $Q = 0$ and so $D = 1$.*

2. The set X is an ESSp set for ${}^N G$, the N times repeated game.

3. σ is a behavior strategy for ${}^N G$ and $\sigma \in X$.

We begin by formalizing the concept of "open minded" mutation. A mutant is open minded when her behavior on the equilibrium path does not deviate from the one of the population but, when facing somebody who deviates, tries to cooperate with her by playing a.

[17] In the examples we will continue to use larger integers, to avoid dealing with too many fractions

Definition 5.1. *We say that σ' is an* open minded *mutation of σ if $\sigma'(h) = \sigma(h)$, when $p^\sigma(h) > 0$ and $\sigma'(h) = a$ when $p^\sigma(h) = 0$*

It is obvious that open minded mutants are silent mutants, as we will use often this fact we state it formally:

Proposition 5.1. *Let X be an ESSp and let $\sigma \in X$, let σ' be an open minded mutation of σ, then σ' is a silent mutation of σ and so, by proposition 3.2, $\sigma' \in X$ and ${}^N U(\sigma) = {}^N U(\sigma')$.*

In the following, whenever we will construct a new strategy in X, we will use this proposition to assume that it is open minded.

If we want to escape from an inefficient equilibrium, apart from being kind to newcomers, we should also be able to recognize them ans so we should avoid noisy actions, as we defined here:

Definition 5.2. *We say that a strategy σ is* not babbling *at history h_t if $supp\,(\sigma(h_t)) \neq \mathbf{A}$.*

A history at which σ is not babbling will be called a not-babbling history (for σ). Not-babbling, open minded strategies allow mutants to make themselves recognizable and coordinate optimally with them, so, if they belong to an ESSp set, they must themselves have high payoffs as the next proposition shows:

Proposition 5.2. *Let $\sigma \in X$, X ESSp, and σ open minded, assume that it does not babble at h_t, then $U_{h_t}(\sigma) \geq N - t - 1$.*

Proof: Let $x \notin supp\,(\sigma(h_t))$ and let τ be the elementary mutation that plays x on h_t and coincides with the open minded σ otherwise. When playing against σ and conditional on h_t, τ plays x at step $t+1$, earning at least zero. At this point σ will realize that τ is not another "sigma" and, being open minded, will play a from next step, $t+2$ on, so that ${}^N U_{h_t}(\tau, \sigma) \geq N - t - 1$. Since ${}^N U_{h_t}(\sigma, \sigma) \geq {}^N U_{h_t}(\tau, \sigma)$ by the definition of ESSp set, the result follows. Note that we do not need $p^\sigma(h_t) \neq 0$; if $p^\sigma(h_t) = 0$, we have the stronger inequality $U_{h_t}(\sigma) \geq N - t$ because of open mindedness. ■

By now we have already proved proposition 4.2: just apply proposition 5.2 to $t = 0$.

5.2 Don't babble

Now things become harder: the simple argument of the preceding section fails if the strategy prescribes full support at the early stages of the game. This makes mutants unrecognizable, and prevents them from being able to drive populations out.

We give an example of how such a strategy may look like.

Consider the stage game, of the Stag Hunting type:

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 10 & 6 \\ 9 & 7 \end{pmatrix} \end{array} \tag{5.3}$$

and repeat it two times. Let σ be defined as follows: in the first stage it plays $1/2C + 1/2D$, in the second one it plays C on asymmetric histories (namely (C, D) and (D, C)) and D on the symmetric ones ((C, C) and (D, D)). This strategy is an isolated Nash equilibrium and the singleton $\{\sigma\}$ is an ESSet. In fact if τ is a best reply to it, at time 2 strategy τ must coincide with σ, because the unique best replies to C and D respectively are themselves. At stage one the forward game, defined in appendix A, is

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 17 & 16 \\ 19 & 14 \end{pmatrix} \end{array} \tag{5.4}$$

It is now of the Hawk-Dove type: if τ is doing anything different from $1/2C + 1/2D$, it will earn against itself less than what σ does. So we have ${}^{N}U(\tau, \tau) < {}^{N}U(\sigma, \tau)$, for all best replies τ to σ different from σ: we have no choice but to live with our expected payoff of 16.5 [18].

The strategy σ above is stoutly staunchly stubbornly stable: it is an evolutionary stable strategy, it is an ESSet, it is an asymptotically stable limit point for every sensible dynamics ... yes, but only if we restrict ourselves to *symmetric* strategies. If we allow asymmetric ones its index will become -1 instead of 1 and the strategy will become evolutionary unstable, in every sense you want, see [7]; the two stable, in every sense, outcomes will be (C, D) and (D, C).

[18] In this simple example the payoff is not too bad compared to the optimal one of 20, if you want to see a dramatic loss in payoffs take the more complicated game (4.7).

And this is what can save us in longer games: if the game is just a segment of a longer one and has been preceded by an asymmetric outcome we have a chance. Suppose that an asymmetric history h, say playing $(x, y), x \neq y$ has been observed. A strategy restricting to the one as before on this subgame cannot be in an ESSet not even in an ESSp as we show now.

Take a mutant τ that uses the history as an anticoordination device: if it observes (x, y) it plays, say, C, if it sees (y, x) it plays D, elsewhere it agrees with σ. Assume that the history has non zero probability $p^\sigma(h) = p^\sigma(\bar{h}) \neq 0$.

The mutation τ is obviously elementary.

Now, τ is a best reply to σ because its actions are in the support of σ. Moreover, upon meeting itself after either of the two asymmetric histories, τ has better payoffs than the one σ against τ would get. This because it anticoordinates with itself optimally while σ randomizes. In fact, if $h_t = (x, y)$, we have, $U_{h_t}(\tau, \tau) = 16 \geq U_{h_t}(\sigma, \tau) = 15$, and $U_{\bar{h_t}}(\tau, \tau) = 19 \geq U_{\bar{h_t}}(\sigma, \tau) = 18$. Adding the two contributions, we get $U(\tau, \tau) - U(\sigma, \tau) = p(h)[(16 + 19) - (15 + 18)] = p(h)[35 - 33] > 0$. So σ is not in an ESSet, nor in a ESSp, any more.

There is another necessary ingredient in our proofs: suppose at some point we hit a forward game such as

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \end{array} \tag{5.5}$$

and that $\sigma(h)$ is $1/2C + 1/2D$, earning $1/2$ against itself. "‘Well, the case seems to be easier than the preceding one, we take as our elementary mutation $\tau(h) = C$, and continue our proof with it ..."’ Wrong!: if we do so we get indeed a new τ in the ESSp set X but we have raised its payoffs to one, so, if we want a lower bound on ALL payoffs in the ESSp this τ is useless. The right trick here is to remember that what we look for are strategies that, at the beginning of the game, are better in sending messages, rather than getting high payoffs. Actually we will choose, among the possible elementary mutation of σ the one with *lowest* payoff and this is why we introduced the concept of submutation, defined in 3.6.

Proposition 5.3. *Let $\sigma \in X$, X ESS, let h_t be an asymmetric history, then there is an elementary submutation of σ at h_t, σ', such that both $\sigma'(h_t)$ and $\sigma'(\bar{h}_t)$ are pure actions.*

In particular, the new strategy, σ', will not babble, $\sigma' \in X$ and ${}^N U(\sigma', \sigma') \leq {}^N U(\sigma, \sigma)$. The proof, elementary but a little technical, is given in the Appendix.

Iterating this procedure we get:

Proposition 5.4. *Given an X ESSp and a $\sigma \in X$ there is a $\sigma' \in X$ such that ${}^N U(\sigma') \leq {}^N U(\sigma)$ and ${}^N U_{h_t}(\sigma') \geq N-t-1$, if h_t is an asymmetric story.*

Proof: It is straightforward: we chose some ordering of the asymmetric histories, starting from $t = 1$ on, so that longer histories come after the shorter ones. Then we perform the submutations of prop 5.3 on non zero probability couples of asymmetric histories, couple by couple in the chosen order.

At each step the strategy becomes a pure action one on a couple, payoffs do not increase and the strategy on the rest of the histories does not change [19].

We repeat the process until we get a strategy that is not-babbling at all nonzero probability asymmetric histories, is still in X and has payoffs not higher than the original one. ■.

If needed, a sequence of silent mutations on zero probability histories, as in subsection 5.1, makes the σ' of the preceding proposition open minded, too. All this without leaving X.

So, if we call σ again the new strategy we constructed, we can recapitulate its properties:

Proposition 5.5. *Given an X, ESSp for ${}^N G$ there is a $\sigma \in X$ such that:*

1. *Conditional on zero probability histories, it would play the optimal strategy, i.e if $p^\sigma(h_t) = 0$ then $\sigma'(h_t) = a$, and so ${}^N U_{h_t}(\sigma) \geq N - t$.*
2. *Conditional on no-babbling histories it earns the efficient symmetric payoff up to one, i.e. if supp$(\sigma'(h_t)) \neq \mathbf{A}$ then ${}^N U_{h_t}(\sigma) \geq N - t - 1$.*
3. *Conditional on asymmetric histories it earns the efficient symmetric payoff up to one i.e. if $h_t \neq \bar{h}_t$ then ${}^N U_{h_t}(\sigma) \geq N - t - 1$.*
4. *For any other $\sigma' \in X$, we have ${}^N U(\sigma', \sigma') \geq {}^N U(\sigma, \sigma)$.*

[19]Choosing the ordering to be compatible with the length is not really necessary here, it has the advantage that it makes the process faster, because all but one descendants of an asymmetric history get probability zero. Moreover, in the case of infinitely repeated games treated later, it is the natural thing to do.

Points 1 to 3 have already been proved, point 4 follows by taking a strategy in X with lowest payoff, this exists because X is compact, and then performing the silent mutation and the submutation of this section on it.

Of course, in points 2 and 3, we do not know exactly what σ plays, we just know what it earns.

5.3 Be patient

We have proved that, if σ is as in proposition 5.5, it will start getting the symmetric efficient payoff as soon as a non-babbling or asymmetric history obtains. To finish the proof we will proceed as illustrated in the example given below. Namely if a strategy prescribes actions with full support at every stage, the probability of an asymmetric outcome is non zero and, after a certain number of stages, symmetric histories will have very low probability and so their contribution to the expected payoff will be irrelevant compared to the high paying asymmetric ones.

There is however a problem to be solved first. If the strategy prescribes an almost pure action, i.e. one that gives probability almost one to a pure action, an asymmetric outcome will be possible but unlikely. So if the mixed actions become more and more pure when the game progresses, it will take a long time before enough asymmetric histories appear and they won't have enough time to get a significant payoff.

This problem is dealt with in proposition 5.6, that shows that if the conditional payoff of a history is low our σ must be enough mixing.

The next subsection 5.4 completes the proof along the lines of our example. An appropriate k is guessed, it must be large enough to allow the probability of symmetric histories to dwindle but not too large so that asymmetric histories have the time to hoard high payoffs in the rest of the game. A good choice is the square root of N, that gives our estimate on $c(N)$. To estimate the rate at which asymmetric histories before k are produced we use proposition 5.6 together with the optional stopping theorem.

The role of the choice of k is illustrated in the following example Consider the stage game

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array} \tag{5.6}$$

repeat it N times, N large. Let σ be the strategy that begins by playing

$1/2C + 1/2D$ and continues to do so until an asymmetric outcome is realized, at this point it repeats the last move.

Chose a t. The probability that the history, after t stages, is still symmetric is $1/2^t$. If at time t an asymmetric outcome is played, players can anticoordinate and get 1 at each of the next stages, so altogether $N - t$. It follows that the payoff of σ is at least $(1 - 1/2^t)(N - t)$.

If we vary t we get ${}^N U(\sigma) \geq \max_{1 \leq t \leq N}(1 - 1/2^t)(N - t)$ so taking e.g. $N = 1032$ and $t = 10$ we get ${}^{10}U(\sigma) \geq 1021 + 1/512$, a rather good estimate.In fact the mistake is less than 1 % given that the true value is little more than 1031 by the following exercise:

Exercise 5.1. *Prove that if* ${}^N G$ *and* σ *are as above,* ${}^N U(\sigma) = N - 1 + 1/2^N$

Hint: Show that the forward game ${}^N G^{\sigma}_{\aleph}$ is

$$\begin{array}{c c} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} {}^{N-1}U(\sigma,\sigma) & N \\ N & {}^{N-1}U(\sigma,\sigma) \end{pmatrix} \end{array} \tag{5.7}$$

and use recursion.

I cheated a little bit because I took the most favorable case: the mixed actions here are always $1/2C + 1/2D$, so the probability of a symmetric history decreases very fast: it gets halved at each step and it is easily seen that the optimal t is of the order of $logN$, in words, very few step are spent in communicating and forging a correlation device to play efficiently.

In general life is not so easy: the mixed actions are not constant, so *a priori* it might happen that the probability of generating an asymmetric action goes to zero so fast that bad asymmetric histories survive for a long time.

The goal of the next set of propositions is to prove that this is not the case. The reason is intuitively simple. If a mixed action played independently on both sides has a low probability of generating an asymmetric outcome, it means that there is a pure action, say x that will be played with probability almost one.

But if the payoff of the strategy is low and the strategy satisfies proposition 5.5 this is impossible: indeed, playing against the strategy a pure action y, different from x,would generate an asymmetric history with high payoff, so the original action would not be a Nash equilibrium in the forward game and σ could not be in an ESSp set, by exercise A.4 in the appendix.

This is what the next proposition formally states, recall that $\sigma(h_t)(x)$ is the probability that, after history h_t, σ takes action x.

Proposition 5.6. *Let $\sigma \in X$, let $p^\sigma(h_t) > 0$, supp$(\sigma(h_t)) = \mathbf{A}$ and h_t symmetric, let σ satisfy proposition 5.5.*

If $U_{h_t}(\sigma) \leq N - A$ then $\forall x \in \mathbf{A}$, $\sigma(h_t)(x) \leq 1 - \frac{A-t-2}{N-t-2}$ and the probability that the outcome is asymmetric is at least $\frac{A-t-2}{N-t-2}$.

Proof: Let x be given and let $y \neq x$ be another action. Let us consider the forward game G_{h_t}.

We have $V^\sigma_{1,h_t}(y, x) \geq (N - t - 2)$ because after (x, y) the history will be asymmetric and σ satisfies 5.5. So, given that payoffs are linear and nonnegative, $V^\sigma_{1,h_t}(y, \sigma) \geq (N - t - 2) \cdot \sigma(h_t)(x)$.

On the other side, action $\sigma(h_t)$ is a best reply to itself in the forward game so $V^\sigma_{1,h_t}(y, \sigma(h_t)) \leq V^\sigma_{1,h_t}(\sigma(h_t), \sigma(h_t)) = U_{h_t}(\sigma(h_t)) \leq N - A$, remember that h_t is symmetric.

Putting together the inequalities you get the bound on $\sigma(h_t)(x)$. The lower bound on the probability of the outcome follows from the next lemma. ■

Lemma 5.1. *If in a stage game every action is taken with probability at most $1 - \delta$, $0 \leq delta \leq 1/2$, the outcome is asymmetric with probability at least δ.*

Proof: If $p_1, p_2, \ldots, p_n$ are the probabilities of actions $a_1, a_2, \ldots, a_n$ respectively, the probability of an asymmetric outcome will be $1 - \sum p_i^2$. It is a concave function and we look for its minimum on the convex set $\mathbf{L} = \{(p_1, p_2, \ldots, p_n) | 0 \leq p_i \leq 1 - \delta, \sum p_i = 1\}$.It is achieved on one of the extreme points that , up to permutation of indexes, is $(1 - \delta, \delta, 0, \ldots, 0)$. The value there is $2\delta(1 - \delta) \geq \delta$. ■

5.4 Completion of the proof

We prove now our bound on the inefficiency $c(N)$ when N goes to infinity. To do so we will concentrate on the first periods of the game, when the role of action as message carriers is more pronounced.

Suppose a σ as in proposition 5.5 has been taken and is fixed in the rest of the proof. We claim that σ has payoff at least $N - c(N)$, with $c(N)$ as in proposition 4.1. Choose a small $\varepsilon > 0$ and let $\beta = 1/2 + \varepsilon$. We call a history "bad" if $U_{h_t}(\sigma) < N - N^\beta$, if not it will be called good. Of course the distinction is useful only for $t \leq N^\beta$: after that time all histories are bad.

If $t \leq N^{1/2} - 1$ and if h_t is any asymmetric history we know, by assumption 5.5, that $U_{h_t}(\sigma) \geq N - N^{1/2}$. So for t up to $N^{1/2} - 1$ only symmetric histories can be bad.

Note that, because of formula (A.20), a bad history can generate a good one and a good history con generate a bad one. On the other side, once a history has become asymmetric, all its descendants will be so.

Let us now fix $k = N^{1/2} - 1$ and consider histories generated by (σ, σ) at time k. A look at figure 1 may help the reader to follow the arguments.

We will partition histories in disjoint classes.

One class, named A_k, will contain all the asymmetric histories.

As for the symmetric ones, some of them can be good or can be bad but have good ancestors. Given such a one, h_k, we denote by $g(h_k)$ its *oldest* good ancestor, this of course could be h_k itself. We will call S_k the set of all oldest ancestors: S_k contains some symmetric histories g_t with $t \leq k$. For each $g_t \in S$, we will denote by $D_k(g_t)$ the set of its symmetric descendants at time k,it can contain both good or bad histories. The $D_k(g_t)$ form a collection of disjoint sets, whose union will be called B_k.

Symmetric histories in the complement of B_k are bad and all their ancestors are, they will be called consistently bad. This set will be denoted by C_k.

To resume, given a stage game G, an N, and a strategy satisfying assumptions 5.1 and 5.5, the histories generated by it at time k are divided in:

1. The set of asymmetric ones A_k
2. Sets of symmetric good ones or symmetric bad ones with a good ancestor, grouped according to the first good ancestor $D_k(g_t)$, $g_t \in S_k$.
3. The set of consistently bad ones, C_k

To prove that ${}^N U(\sigma)$ is large for long games we show first that the bad set C_k is small for N large. We denote by $p_N(C_k)$ the probability of event C_k in a N times repetition of the game: it is a function of N and also of σ and G.

In the next proposition we show not only that that it is small for N large but also that the upper bound does not depend on σ or G, provided assumption 5.1 is satisfied :

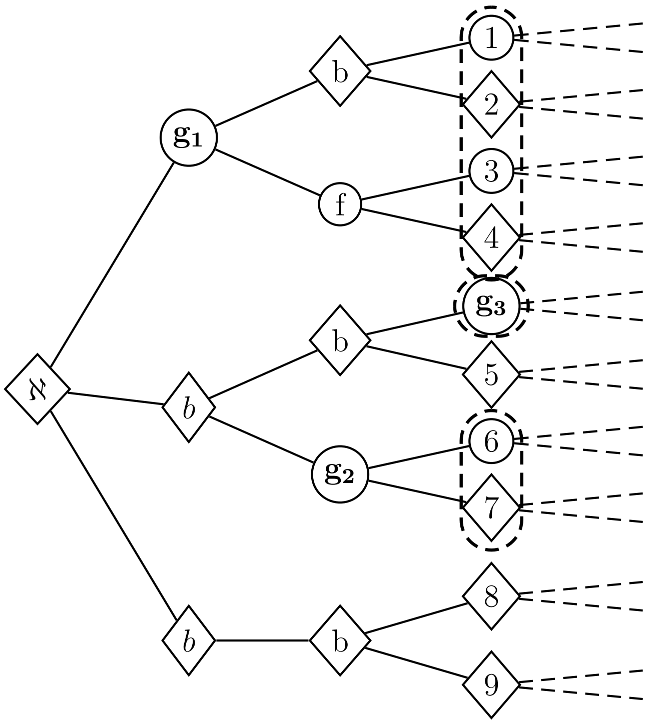

Figure 1: A possible tree. Round circles are good symmetric histories, diamond ones are bad symmetric, asymmetric ones are not shown. At time 3, class B consists of histories 1,2,3,4,in the equivalence class $D_3(g_1)$, 6 and 7 in $D_3(g_2)$,and g_3 in $D_3(g_3)$. Class C contains the consistently bad histories 5, 8 and 9. So we have that ${}^N U(\sigma) \geq \sum_{h_3 \text{ asymmetric}} {}^N U_{h_3} + p^\sigma(g_1)^N U_{g_1} + p^\sigma(g_2)^N U_{g_2} + p^\sigma(g_3)^N U_{g_3} + \text{bad histories}$.

Lemma 5.2. *Let ${}^N G$ and σ satisfy 5.1 and 5.5, and let ε and k be as above, then $p_N(C_k)$ decreases exponentially when N goes to infinity, more precisely there is a $\bar{N}$ such that for $N \geq \bar{N}$ we have $p_N(C_k) \leq exp(-N^\varepsilon/6)$. Moreover $\bar{N} = \max(4, N_1, 2^{1/\varepsilon})$, where N_1 is a universal constant, independent of ε.*

Proof: By propositions 5.6 if h_t is bad, it will generate an asymmetric history at stage $t+1$ with probability at least $\frac{N^\beta - t - 2}{N - t - 2}$, so the probability measure of C_k will be bounded by $\prod_{t=0}^{t=k-1}\left(1 - \frac{N^\beta - t - 2}{N - t - 2}\right)$, with $k = N^{1/2} - 1$. The estimate of this expression is a calculus exercise that can be found in the appendix. ∎

We can now conclude the proof. The evaluation of ${}^N U(\sigma)$ is very quick if we use the fact that $U_{h_t}(\sigma)$ is a supermartingale.

We define a stopping time θ as

$$\theta = \min\left(\inf\{t | h_t \text{ is good}\}, k\right)$$

Se we have that either h_θ is a good (symmetric or asymmetric) strategy with $U_{h_\theta}(\sigma) \geq N - N^\beta$ or that $\theta = k$ and h_k is consistently bad, the latter case will happen with probability $p_N(C_k)$.

Now use (A.21):

$$^NU(\sigma) \geq E(^NU_{h_\theta}) \geq \big(1 - p_N(C_k)\big)(N - N^\beta) + \sum_{C_k} p(h_k)^N U_{h_k} \tag{5.8}$$

The sum over C_k contributes at least zero and so by our estimate of $p_N(C_k)$, lemma 5.2, $^NU(\sigma) \geq (N - N^\beta)(1 - exp(-N^\varepsilon/6)$ for $N \geq \bar{N}$.

If you do not like supermartingales you can prove the same inequality directly using the partition given above and equation (A.22). Again a look at picture 1 may help.

So in the end we have

$$^NU(\sigma) \geq U(\sigma') \geq (N - N^\beta)(1 - exp(-N^\varepsilon/6) > N - N^\beta - Nexp(-N^\varepsilon/6) \tag{5.9}$$

for N larger than the $\bar{N}$ given above. More precisely, looking at the expression for $\bar{N}$ we see that inequality 5.9 holds if $1/2 > \varepsilon > \frac{log2}{logN}$ and $N \geq N_1$. We define:

$$c(N) = \begin{cases} \inf_{1/2 > \varepsilon > \frac{log2}{logN}} N^{1/2+\varepsilon} + Nexp(-N^\varepsilon/6) \text{ for } N \geq N_1 \\ N \text{ for} N \leq N_1 \end{cases} \tag{5.10}$$

Now we have $^NU(\sigma) \geq N - c(N)$ and it is an easy exercise to see that $\lim_{N\to\infty} \frac{c(N)}{N^{1/2+\varepsilon}} = 0$ for all ε.

This ends the proof.

6 Infinitely repeated games with discounting

In this section we deal with the case of infinitely repeated games with discounting, we will use the notation for finite ones in section 2.2, apart from the following changes:

Time: The set of time periods will be now all the integers from 0 to ∞ : $\{0, 1, 2, \dots\}$

Perfect monitoring: no change

Histories: Histories will now be:

$$H = \{\aleph\} \cup \bigcup_{t=0}^{\infty} H_t \tag{6.1}$$

so the union of a one element set, now the -1 history, $h_{-1} = \aleph$, and the $H_t = (\mathbf{A^2})^{t+1}$ for each $t \geq 0$. So H_t will now be the set of set of $[(a_0, b_0), (a_1, b_1), \dots, (a_t, b_t)]$.

Strategies: no change. But now they form an infinite dimensional space, so we will need to give a topology for them, this is done in subsection 6.2.

Plays: no change

Probabilities: no change.

Payoffs: The payoffs will now be discounted at rate ρ: so the payoff of strategy σ against strategy τ, denoted by ${}^{\rho}U(\sigma, \tau)$, will be :

$${}^{\rho}U(\sigma, \tau) = (1 - \rho) E_{(\sigma,\tau)} \left(\sum_{t=0}^{+\infty} \rho^t u(a_t, b_t) \right) \tag{6.2}$$

Multiplication by $(1-\rho)$ makes the repeated-game payoffs easily comparable with the stage-game payoffs. In particular, a player who earns constantly the same stage-game payoff u in each period will have repeated-game payoff u. The expectation is taken with respect to the probability measure induced by the strategy profile (σ, τ).

The conditional payoff ${}^{\rho}U_{h_t}(\sigma, \tau)$ is defined as

$${}^{\rho}U_{h_t}(\sigma, \tau) = (1 - \rho) E_{(\sigma,\tau)|h_t} \left(\sum_{s=t+1}^{+\infty} \rho^s u(a_s, b_s) \right) \tag{6.3}$$

Note that we discount the value at time 0.

The forward game is defined as before but now payoffs are

$$ {}^{f}V^{(\sigma,\tau)}_{1,h_t}(x,y) = (1-\rho)\rho^t u(x,y) + {}^{N}U_{h_t\circ(x,y)}(\sigma,\tau) \tag{6.4} $$
$$ {}^{f}V^{(\sigma,\tau)}_{2,h_t}(x,y) = (1-\rho)\rho^t u(y,x) + {}^{N}U_{\bar{h}_t\circ(y,x)}(\tau,\sigma) \tag{6.5} $$

We will be interested in the behavior of ${}^{\rho}G$ when ρ is close to one. It is well known that ${}^{\rho}G$ can be interpreted as a game in which future payoffs are undiscounted, but the probability that the game continues to the next stage is ρ, in this case, the expected length of the game $\bar{N}$ will be $(1-\rho)^{-1}$, or $\rho = \big(1 - 1/\bar{N}\big)$, this fact will be used to compare with the results for N times repeated games.

6.1 Results

Proposition 6.1. *There is a universal function $f(\rho)$ that goes to* 1 *when ρ goes to* 1 *such that: if G is a symmetric game with P and Q as before, ${}^{\rho}G$ is the corresponding discounted game and σ is a strategy in an ESSp set of ${}^{\rho}G$:*

$$ {}^{\rho}U(\sigma) \geq f(\rho)P + (1 - f(\rho))\,Q \tag{6.6} $$

In the proof it will be seen that we can take $f(\rho) = \rho^{(-log\rho)^{-(1/2+\varepsilon)}}$, for ρ close to one. It is trivial that this function goes to 1 when ρ goes to 1. More interesting is to see how the estimate looks like when we write the discount factor as $\rho = (1 - 1/\bar{N})$, $\bar{N}$ the expected number of rounds. Then $f(\rho) \approx \rho^{\bar{N}^{(1/2+\varepsilon)}}$. The interpretation should be clear: as in the case of finitely repeated games a certain number of rounds, of order of magnitude $\bar{N}^{(1/2+\varepsilon)}$, are needed to convey understand and implement the message: "Let's play the optimal action instead of this nonsense".

Theorem 3. *Let G be a Paretian Game with optimal action a, optimal payoff P and worse payoff Q and consider the infinitely repeated repeated game with discount ${}^{\rho}G$. Then*

1. *The set $\{\sigma | {}^{\rho}U(\sigma) = N \cdot P\} = \{\sigma | \sigma(h_t) = a \text{ if } p^{\sigma}(h_t) > 0\}$ consisting of strategies playing always a against themselves is an ESS.*
2. *If $\sigma \in X$, X an ESS, then $lim_{\rho\to 1}{}^{\rho}U(\sigma) = P$*

The proof is as in the finitely repeated case.

6.2 Proofs

The proofs are similar or easier. We will just point to the few technicalities involved and leave the rest to the reader as an exercise.

First we have to specify the choice of topology for for the infinite dimensional space ${}^{\delta}\mathcal{B}$.

Strategies in ${}^{\delta}\mathcal{B}$ are functions from the set of histories to the simplex of mixed actions. We will give to it the standard product topology: given a finite set of histories S and a $\delta > 0$, a neighborhood of a strategy σ will be the set of those strategies that, on histories in S, take mixed actions within δ of the one taken by σ. In practice this means that a sequence of strategies σ_i converges to σ if and only if, for each history h_t, $lim_{i\to\infty}\sigma_i(h_t) = \sigma(h_t)$, not necessarily uniformly in h_t.

The reader should check, it is a standard exercise, that the payoff function (6.2) is continuous with this topology.

We can now use the definition of ESSet given in 3.2, note that the requirement that X is closed is crucial, we had to give ${}^{\delta}\mathcal{B}$ a topology because of this.

Proposition 3.2 applies, not that now the mutation takes place at infinitely many histories, so apart from induction, closedness is needed. The reader can check it as a standard exercise.

Proposition 5.1 is the same. Again we need closure

Proposition 5.2 has bound ${}^{\rho}U_{h_t} \geq \rho^{t+2}$.

Proposition B.1 is the same.

Proposition 5.4 has bound ${}^{\rho}U_{h_t} \geq \rho^{t+2}$

Proposition 5.5 is the same, provided we change the values of the efficient payoffs from $N - t$ to ρ^t and so on.

Proposition 5.6 is now as follows.

Proposition 6.2. *Let $\sigma \in X$, let $p(h_t) > 0$, $supp(\sigma(h_t)) = \mathbf{A}$ and h_t symmetric, let σ be as in proposition 5.5. If $U_{h_t}(\sigma(h_t)) \leq \rho^{\alpha}$ then $\forall x \in \mathbf{A}$ $\sigma(h_t)(x) \leq \rho^{\alpha-t-3}$*

Proof: Given x, let $y \neq x$ the utility of playing y against $\sigma(h_t)$ in the forward game is less than ρ^{α}, the utility of playing y against x is at least ρ^{t+3} by proposition 5.4. So we have $\sigma(h_t)(x)\rho^{t+3} \leq \rho^{\alpha}$ and the result follows. ■

Now set $\alpha = (-log\rho)^{-(1/2+\varepsilon)}$ and $\beta = integerpart(-log\rho)^{-1/2}$. This time bad symmetric histories will be those for which $U_{h_t}(\sigma(h_t)) \leq \rho^{\alpha}$. It is easily seen that, when ρ goes to one, α, β and α/β go to to ∞. In particular $\rho^{\beta+3} > \rho^{\alpha}$ when ρ is close enough to 1 and so histories that become asymmetric before time β will be good.

The probability that h_{t+1} is bad and symmetric conditional on h_t being bad and symmetric is at most $\rho^{\alpha-t-3}$ by proposition 6.2. So consistently bad histories from $t = 0$ to $t = integerpart(-log\rho)^{-1/2}$ have probability at most $\prod_{t=0}^{t=beta-1} \rho^{\alpha-t-3} = \rho^{\beta^2 \frac{(\alpha-3)}{\beta} - 1/2(1-1/\beta)}$. By substitution of the values of α and β it is easily seen that this goes to zero when ρ goes to one. So at time β most histories are good and the result follows.

7 Behavioral Maxims

In this section we deal with a technically easy but, in the author's opinion, conceptually important extension of our results. Suppose that the game changes at each stage and that it even depends on the previous history. Still, we require that some gross features of it are preserved. All stage games must be paretian, up to a permutation the optimal action may be assumed to be a, the payoff of a against itself is at least P and the cost of miscoordination is at most D. The optimal strategies may be complicated, involving a trade off between current payments and histories leading to high paying games, still the general principle holds: evolution leads to average payoff per stage at least P.

More formally, given $\mathbf{A}$ and $a \in \mathbf{A}$, P and D, let $\mathcal{G}(P, D)$ be the set of all payoffs functions $u : \mathbf{A}^2 \to \mathbf{R}$ such that

1. For all $u \in \mathcal{G}(P, Q)$ we have: $a \in \operatorname{Argmax}_{x \in \mathbf{A}} u(x, x)$ and $u(a, a) \geq P$
2. For all $u \in \mathcal{G}(P, Q)$ we have: $P - \min_{x,y \in \mathbf{A}} u(x, y) \leq D$

They define the class of symmetric games described informally above. Let $u_h : H \to \mathcal{G}(P, D)$ be a function from histories to them. It defines a *iterated game* ${}^N\tilde{G}$ that has the same times, histories and strategies as that in section 2.2, and where the new payoffs are now defined as

$$ {}^N\tilde{U}(r_N) = \sum_{i=1}^{N} u_{r_i}(a_i, b_i) \tag{7.1} $$

where r_i is the first i segment of history r_N.

The following theorem holds:

Proposition 7.1. *There is a universal function* $c(N)$ *with* $\lim_{N\to\infty} \frac{c(N)}{N^{1/2+\varepsilon}} = 0$ *for all* $\varepsilon > 0$*, such that:*

if ${}^N\tilde{G}$*,* P *and* D *are as above and if* $\Pi(N) = \inf\left\{{}^N U(\sigma)|\sigma \in X,\ X\ ESSet for {}^N\tilde{G}\right\}$ *then:*

$$\Pi_N \geq P \cdot N - c(N) \cdot D \tag{7.2}$$

The proof is essentially the same as that of proposition 4.1. It can be improved taking into account the above mentioned trade off between current and future payoffs. Statement and proof are left to the reader.

8 Conclusions

It was proved that in long repeated games, the evolutionary stable payoffs are asymptotically efficient. The results in this paper can be extended in several directions. A similar one holds for asymmetric games and for multiplayers games, we leave to the reader a formulation and proof of it.
More interesting is to see what happens in games such as the, finitely repeated, Hawk Dove or the Prisoner's Dilemma game, in which new phenomena occurs. Here too evolution leads from inefficient to efficient strategies according to the same mechanism as the one described in this paper. The problem is that it does not settle there and it should not because there are no ESSets. If they existed, on one side they should be asymptotically efficient, because of proposition 4.1, on the other one they would consist of Nash equilibria, but the only equilibria in the finitely repeated prisoner's dilemma are inefficient [20].

We are not happy with just a negative result and intuition suggests what a satisfactory extension should be. In these cases, it is not only enough to be "'kind to foreigners"' but, in order to survive, one should also be able to "'deter those who exploit my kindness"'.

In the setting of payoff consistent dynamics, mentioned in the introduction, this corresponds to an attractor that contains non Nash equilibria, evolution approaches efficient strategies but then, to account of the possibility of "'evil"', and stupid, mutants, leads astray from it. It is interesting to see what happens if a continuous stream of mutants is assumed, this is

[20]In the discounted case there are efficient components of NE, but they can be proved directly to be vulnerable even to elementary mutations

an idea that was already implicit in the original work of Maskin and Fudenberg [12]. The minimal asymptotically stable sets shrink to sets of strategies that cooperate but, upon defection, are able to react switching to punishment. The more sophisticated the entering mutants are assumed to be, say experimenting a one time defection and then being open to cooperation, the smaller the minimal asymptotically stable set will be allowing more and more refined strategies see [3] for some analogous ideas. This will be the topic of the following paper.

What are the new beh rules? intuition suggests something like "Be kind to newcomers, unless they cheat"', "'Punish but to a moderate extent"' etc.

References

[1] R. Aumann, Nash equilibria are not self enforcing, IJGT.

[2] J.P. Benoit, V. Krishna, Finitely Repeated Games, Econometrica, Vol. 53, No. 4 (Jul., 1985), pp. 905-922.

[3] K G. Binmore, L. Samuelson, Evolutionary Stability in Repeated Games Played by Finite Automata, JET 57, 278-305 (1992).

[4] Cima A., F. Mañosas and J. Villadelprat, A Poincaré-Hopf theoremfor noncompact manifolds, *Topology*, **37** (1998), 261-277.

[5] S. Demichelis, On several Misconceptions on Evolution and Rationality in Game Theory, in preparation.

[6] S. Demichelis, Efficient Communication and Repetition: an Evolutionary Foundation of Conversational Maxims, in preparation.

[7] S.Demichelis and K.Ritzberger, From evolutionary to strategic stability, *Journal of Economic Theory*, (2003), to appear.

[8] S.Demichelis, K. Ritzberger, and J.Swinkels, The SimpleGeometry of Perfect Information Games, IHS WorkingPaper 115, (2002), *International Journal of Game Theory*, to appear.

[9] S. Demichelis, J. Weibull, "Language, Meaning, and Games: A Model of Communication, Coordination, and Evolution," American Economic Review, vol. 98(4), pages 1292-1311, September.

[10] Dreber A, DG Rand, D Fudenberg, MA Nowak (2008). Winners don't punish. Nature 452: 348-351

[11] R. Durrett: Brownian motion and martingales in analysis,Wadsworth Pub. Co.

[12] D. Fudenberg, E. Maskin (1990):"Evolution and Cooperation in Noisy Repeated Games", American Economic Review, 80, 274-279.

[13] D. Fudenberg, E. Maskin (2007):"Evolution and Repeated Games", .

[14] P. Grice, "Logic and conversation". In Cole, P. and Morgan, J. (eds.) Syntax and semantics, vol 3. New York: Academic Press,1975.

[15] I.Karatzas, S. Shreve ,Browniam motion and Stochastic calculus, Springer

[16] Y. Kim and J.Sobel, An Evolutionary Approach to Pre-Play Communication, Econometrica, Vol. 63, No. 5 (Sep., 1995), pp. 1181-1193

[17] S. Levinson, Pragmatics, Cambridge University Texts, 1983.

[18] G. J. Mailath, L. Samuelson, *Repeated Games and Reputations*, Oxford, (2006).

[19] J. Maynard Smith, Price, G.R. (1973). "The logic of animal conflict". Nature 246 (5427): 1518.

[20] B. Thomas (1985):On evolutionarily stable sets, J. Math. Biology 22, 105-115

[21] D. Revuz, M.Yor, Continuous Martingales and Brownian Motion, Springer

A Additional notation and tools

In this technical part we describe a basic tool for many of our proofs: the forward game. We assume that N has been fixed and so, when no risk of confusion arises, we will disregard it in our notation.

Definition A.1. *The* forward game ${}^{f}G_{h_t}^{(\sigma,\tau)}$ *is a two player game* $(\mathbf{A}, \mathbf{A}, {}^{f}V_{1,h_t}^{(\sigma,\tau)}, {}^{f}V_{2,h_t}^{(\sigma,\tau)})$ *in which the pure strategy set consist of the set of actions* $\mathbf{A}$ *for both players and, given an action pair* $(x, y) \in \mathbf{A}$ *the payoffs are:*

$$ {}^{f}V_{1,h_t}^{(\sigma,\tau)}(x, y) = u(x, y) + {}^{N}U_{h_t \circ (x,y)}(\sigma, \tau) \tag{A.1} $$

$$ {}^{f}V_{2,h_t}^{(\sigma,\tau)}(x, y) = u(y, x) + {}^{N}U_{\bar{h}_t \circ (y,x)}(\tau, \sigma) \tag{A.2} $$

Payoffs are extended to mixed actions in the usual way.

The interpretation of this game is the following: player 1 is a player that has observed history h_t and chooses action x , player 2, who has observed history $\bar{h}_t$, choses y. To the payoffs of the stage game we then add what the players would get if they used the relevant parts of strategies σ and τ applied to histories $h_t \circ (x, y)$ and $\bar{h}_t \circ (y, x)$ and to their descendants from time $t+2$ on.

We stress the fact that ${}^{f}G_{h_t}^{(\sigma,\tau)}$ is a one stage game, the only choice is (x, y), then the behavior is fixed by σ and τ.

We give below some properties of forward games to be used in our study of mutations. First note that, if $\sigma = \tau$ *and* $h_t = \bar{h}_t$, the game is symmetric. In general we have ${}^{f}V_{1,h_t}^{(\sigma,\tau)}(x, y) = {}^{f}V_{2,\bar{h}_t}^{(\tau,\sigma)}(y, x)$.

Strategies pairs (σ, τ) in ${}^{N}G$ induce strategy pairs $\big(\sigma(h_t), \tau(\bar{h}_t)\big)$ for player 1 and 2 in all games ${}^{f}G_{h_t}^{(\sigma,\tau)}$ and we have:

$$ {}^{N}U_{h_t}(\sigma, \tau) = {}^{f}V_{1,h_t}^{(\sigma,\tau)}(\sigma(h_t), \tau(\bar{h}_t)) \tag{A.3} $$

and

$$ {}^{N}U_{h_t}(\tau, \sigma) = {}^{f}V_{2,h_t}^{(\sigma,\tau)}(\sigma(h_t), \tau(\bar{h}_t)) \tag{A.4} $$

In the proof of Theorem 2 in subsection 4.1 we will also need the *continuation game* that we define below following [18]. In it, unlike the forward game, every action after time $t + 1$ may be chosen.

Definition A.2. *The* continuation game ${}^{c}U_{h_t}$ *is the* $N - t$ *repeated game* ${}^{N-t}G$ *beginning at time* $t + 1$. *A strategy* σ *in* ${}^{N}G$ *induces, by restriction to descendants of* h_t, *a continuation strategy denoted by* ${}^{c}\sigma_{h_t}$*:* ${}^{c}\sigma_{h_t}(k_{t+1,s}) = \sigma(h_t \circ k_{t+1,s})$.

A.1 Some examples

This section is for reader who want to develop some familiarity with the concept of forward game.

Consider the stage game

$$\begin{array}{cc} & \begin{array}{cc} H & D \end{array} \\ \begin{array}{c} H \\ D \end{array} & \begin{pmatrix} -1 & 5 \\ 3 & 1 \end{pmatrix} \end{array} \tag{A.5}$$

Repeat it two times. Let σ be defined as follows:

$$\sigma(h_t) = \begin{cases} 1/2H + 1/2D & \text{if} h_t = \aleph, (H,H)\text{or}(D,D) \\ H & \text{if} h_t = (H,D) \\ D & \text{if} h_t = (D,H) \end{cases} \tag{A.6}$$

In words: σ plays $1/2H + 1/2D$ at time one and at time two after a symmetric history. If the outcome is asymmetrical, players play again what they have just played.

The game ${}^{f}G^{\sigma}_{\aleph}$ is symmetric and given by

$$\begin{array}{cc} & \begin{array}{cc} H & D \end{array} \\ \begin{array}{c} H \\ D \end{array} & \begin{pmatrix} -1+2 & 5+5 \\ 3+3 & 1+2 \end{pmatrix} \end{array} \tag{A.7}$$

The two added to the diagonal is the payoff of the mixed strategy.

Let now the game be repeated three times and let σ be similarly defined: it begins by playing $1/2H + 1/2D$ and repeats it until an asymmetric history appears, upon an asymmetric history it repeats the last action until the end.

$$\sigma(h_t) = \begin{cases} 1/2H + 1/2D & \text{if} h_t = \aleph, (H,H), (D,D), [(H,H),(H,H)], \\ & \qquad [(H,H),(D,D)], [(D,D),(D,D)], \\ & \qquad [(H,H),(D,D)] \text{ or } [(D,D),(H,H)]. \\ H & \text{if} h_t = (H,D), [(H,D),(*,\bullet)] \text{ or } [(\bullet,\bullet),(H,D))] \\ D & \text{if} h_t = (D,H), [(D,H),(*,\bullet)] \text{ or } [(\bullet,\bullet),(D,H))] \end{cases} \tag{A.8}$$

Here $(\bullet,\bullet)$ denotes any symmetric pair of actions and $(*,\bullet)$ denotes any arbitrary pair.

Then ${}^fG^\sigma_{(H,D)}$ is the asymmetric game

$$\begin{array}{cc} & \begin{array}{cc} H & D \end{array} \\ \begin{array}{c} H \\ D \end{array} & \begin{pmatrix} (-1+5,-1+3) & (5+5,3+3) \\ (3+5,5+3) & (1+5,1+3) \end{pmatrix} \end{array}$$

player one adds to the payoffs of period two the 5 she will get in period 3 by playing H against D, player two adds 3, the payoff of D against H.

The next example shows how the structure of the game can vary during the play.Let us start with the Stag Hunting game

$$\begin{array}{cc} & \begin{array}{cc} H & D \end{array} \\ \begin{array}{c} H \\ D \end{array} & \begin{pmatrix} 5 & 2 \\ 4 & 3 \end{pmatrix} \end{array} \tag{A.9}$$

Let the game be repeated two times and let σ be:

$$\sigma(h_t) = \begin{cases} 1/2H + 1/2D & \text{if} h_t = \aleph \\ H & \text{if } h_t = (H,D),(D,H) \\ D & \text{if } h_t = (H,H),(D,D) \end{cases} \tag{A.10}$$

The game ${}^fG^\sigma_\aleph$ is given by

$$\begin{array}{cc} & \begin{array}{cc} H & D \end{array} \\ \begin{array}{c} H \\ D \end{array} & \begin{pmatrix} 8 & 7 \\ 9 & 6 \end{pmatrix} \end{array} \tag{A.11}$$

note that it has become of the Hawk Dove type. In particular, the only symmetric evolutionary stable strategy is the totally mixed one.

A.2 Easy Exercises and Fruitful Facts on the Forward Game

Given α and β mixed actions in $\Delta(\mathbf{A})$, and a asymmetric history h_t, we define:

$$\begin{aligned} {}^NW^{(\sigma)}_{h_t}(\alpha,\beta) &= \\ &= {}^fV^\sigma_{1,h_t}(\alpha,\beta) + {}^fV^\sigma_{1,\bar{h_t}}(\beta,\alpha) - {}^fV^\sigma_{1,h_t}(\sigma(h_t),\beta) - {}^fV^\sigma_{1,\bar{h_t}}(\sigma(\bar{h_t}),\alpha) \\ &= {}^fV^\sigma_{1,h_t}(\alpha,\beta) + {}^fV^\sigma_{2,h_t}(\alpha,\beta) - {}^fV^\sigma_{1,h_t}(\sigma(h_t),\beta) - {}^fV^\sigma_{2,h_t}(\alpha,\sigma(\bar{h_t})) \end{aligned} \tag{A.12}$$

It is a bilinear form in α, β. The V and the W functions will turn out to be useful in the study of mutants payoffs, as the following propositions show. The proofs are straightforward checks of the definitions and are left to the reader, some informal hints are after the statements.

Exercise A.1. *Let σ' differ from σ only at history h_t, and let $\sigma'(h_t) = \alpha$ then*

$$\begin{aligned}&{}^N U(\sigma', \sigma) - {}^N U(\sigma, \sigma) = \\ &\qquad = p^\sigma(h_t)[{}^f V^\sigma_{1,h_t}(\alpha, \sigma(\bar{h}_t)) - {}^f V^\sigma_{1,h_t}(\sigma(h_t), \sigma(\bar{h}_t))]\end{aligned} \tag{A.13}$$

moreover, when h_t is asymmetric i.e. $h_t \neq \bar{h}_t$, we have:

$$\begin{aligned}{}^N U(\sigma', \sigma') - {}^N U(\sigma, \sigma') = \\ = p^\sigma(h_t)[{}^f V^\sigma_{1,h_t}(\alpha, \sigma(\bar{h}_t)) + {}^f V^\sigma_{1,\bar{h}_t}(\sigma(\bar{h}_t), \alpha) + \\ - {}^f V^\sigma_{1,h_t}(\sigma(h_t), \sigma(\bar{h}_t)) - {}^f V^\sigma_{1,\bar{h}_t}(\sigma(\bar{h}_t), \alpha)] = \\ = p^\sigma(h_t)[{}^f V^\sigma_{1,h_t}(\alpha, \sigma(\bar{h}_t)) - {}^f V^\sigma_{1,h_t}(\sigma(h_t), \sigma(\bar{h}_t))] = \\ = p^\sigma(h_t)[{}^f V^\sigma_{1,h_t}(\alpha, \sigma(\bar{h}_t)) - {}^f V^\sigma_{1,h_t}(\sigma(h_t), \sigma(\bar{h}_t))]\end{aligned} \tag{A.14}$$

and when h_t is symmetric i.e. $h_t = \bar{h}_t$ we have:

$$\begin{aligned}&{}^N U(\sigma', \sigma') - {}^N U(\sigma, \sigma') = \\ &\qquad = p^\sigma(h_t)[{}^f V^\sigma_{1,h_t}(\alpha, \alpha) - {}^f V^\sigma_{1,h_t}(\sigma(h_t), \alpha)]\end{aligned} \tag{A.15}$$

Note first that $p^\sigma(h_t) = p^{(\sigma',\sigma)}(h_t)$, because at histories before t, σ and σ' coincide. Moreover ${}^f V^{(\sigma',\sigma)}_{1,h_t} = {}^f V^{(\sigma,\sigma)}_{1,h_t} = {}^f V^{(\sigma',\sigma')}_{1,h_t}$ because at histories after $t+1$, σ and σ' coincide, too. The only difference is at time t in how ${}^f G^{(\sigma,\sigma)}_{h_t}$ is played. Also remember that $p^\sigma(h_t) = p^\sigma(\bar{h}_t)$.

Now the hints: equation (A.13) is proved by remarking that the strategy profiles (σ', σ) and (σ, σ) generate the same histories with the same probabilities on the part of the game tree not following h_t. So, if you are mutant σ' and you play against the population member σ, the only changes with respect to to a σ playing against another σ will occur after you observe h_t (and so your opponent observes $\bar{h}_t$). This occurs with probability $p^\sigma(h_t)$. In the following periods they revert to the prescriptions of strategy σ.

Equations (A.14) and (A.15) are proved similarly. Note that after simplification expression (A.14) reduces to (A.13): in fact when the mutants meet each other, the one who sees $\bar{h}_t$ plays as the population does. In particular these two expressions are both zero when α is a best reply to the population strategy $\sigma(\bar{h}_t)$.

Exercise A.2. *Let h_t be asymmetric and let σ' differ from σ only at histories h_t and $\bar{h}_t$, and let $\sigma'(h_t) = \alpha$ and $\sigma'(\bar{h}_t) = \beta$ then*

$$\begin{aligned}
&{}^{N}U(\sigma', \sigma) - {}^{N}U(\sigma, \sigma) = \\
&= p^{\sigma}(h_t)[{}^{f}V^{\sigma}_{1,h_t}(\alpha, \sigma(\bar{h}_t)) + {}^{f}V^{\sigma}_{1,\bar{h}_t}(\beta, \sigma(h_t)) + \\
&- {}^{f}V^{\sigma}_{1,h_t}(\sigma(h_t), \sigma(\bar{h}_t)) - {}^{f}V^{\sigma}_{1,\bar{h}_t}(\sigma(\bar{h}_t), \sigma(h_t))] \\
&= p^{\sigma}(h_t)[{}^{f}V^{\sigma}_{1,h_t}(\alpha, \sigma(\bar{h}_t)) + {}^{f}V^{\sigma}_{2,h_t}(\sigma(h_t), \beta) + \\
&- {}^{f}V^{\sigma}_{1,h_t}(\sigma(h_t), \sigma(\bar{h}_t)) - {}^{f}V^{\sigma}_{2,h_t}(\sigma(h_t), \sigma(\bar{h}_t))]
\end{aligned} \tag{A.16}$$

and

$$\begin{aligned}
&{}^{N}U(\sigma', \sigma') - {}^{N}U(\sigma, \sigma') = \\
&= p^{\sigma}(h_t)[{}^{f}V^{\sigma}_{1,h_t}(\alpha, \beta) + {}^{f}V^{\sigma}_{1,\bar{h}_t}(\beta, \alpha) + \\
&- {}^{f}V^{\sigma}_{1,h_t}(\sigma(h_t), \beta) - {}^{f}V^{\sigma}_{1,\bar{h}_t}(\sigma(\bar{h}_t), \alpha)] = \\
&= p^{\sigma}(h_t)[{}^{f}V^{\sigma}_{1,h_t}(\alpha, \beta) + {}^{f}V^{\sigma}_{2,h_t}(\alpha, \beta) + \\
&- {}^{f}V^{\sigma}_{1,h_t}(\sigma(h_t), \beta) - {}^{f}V^{\sigma}_{2,h_t}(\alpha, \sigma(\bar{h}_t))] = \\
&= p^{\sigma}(h_t) {}^{N}W^{(\sigma)}_{h_t}(\alpha, \beta)
\end{aligned} \tag{A.17}$$

Note that, when $\alpha = \sigma(h_t)$ or $\beta = \sigma(\bar{h}_t)$, expressions (A.16) and (A.17) coincide with (A.13) and (A.14) respectively, that in their turn, coincide with each other. The following exercise follows immediately from the definitions.

Exercise A.3. *Let t be any fixed time, then*

$$ {}^{N}U(\sigma, \tau) = \sum_{h_t} p^{(\sigma,\tau)}(h_t)[U(h_t) + {}^{N}U_{h_t}(\sigma, \tau)] \tag{A.18}$$

and

$$ {}^{N}U_{h_t}(\sigma, \tau) = u(\sigma(h_t), \tau(\bar{h}_t)) + \sum_{k_{t+1} | h_t \triangleright k_{t+1}} p^{(\sigma,\tau)}_{h_t}(k_{t+1}) {}^{N}U_{k_{t+1}}(\sigma, \tau) \tag{A.19}$$

The forward game is very handy in studying elementary mutations and ESSp sets, as the following exercise shows.

Exercise A.4. *If* (σ) *belongs to a ESSP than for every* h_t *such that* $p^{(\sigma)}(h_t) \neq 0$*,* $(\sigma(h_t), \sigma(\bar{h_t}))$ *is a Nash equilibrium of* ${}^f G_{h_t}^{(\sigma)}$*.*

This is all we will need. The reader can, however, do the following two instructive exercises:

Exercise A.5. *If* (σ, τ) *is a Nash equilibrium of* ${}^N G$ *then, for every* h_t *such that* $p^{(\sigma,\tau)}(h_t) \neq 0$*,* $(\sigma(h_t), \tau(\bar{h_t}))$ *is a Nash equilibrium of* ${}^f G_{h_t}^{(\sigma,\tau)}$*. Find a counterexample to show that the converse does no hold.*

In fact, strategies inducing Nash equilibria in all ${}^f G_{h_t}^{(\sigma,\tau)}$ correspond just to Nash equilibria in the agent normal form of the repeated game. To get an "iff", we need subgame perfection:

Exercise A.6. *The profile* (σ, τ) *is a subgame perfect equilibrium if and only if, for every* h_t*,* $(\sigma(h_t), \tau(h_t))$ *is a Nash equilibrium of* ${}^f G_{h_t}^{(\sigma,\tau)}$*.*

The if part is an easy backward induction argument, the only if is obvious.

Now suppose that G, the stage game, has all its payoffs ≥ 0. In this case equation (A.19) gives:

$$ {}^N U_{h_t}(\sigma, \tau) \geq \sum_{k_{t+1} | h_t \triangleright k_{t+1}} p_{h_t}^{(\sigma,\tau)}(k_{t+1})\, {}^N U_{k_{t+1}}(\sigma, \tau) = E({}^N U_{k_{t+1}}(\sigma, \tau) | h_t) \tag{A.20} $$

the weak inequality comes from the fact that,given our assumption on the stage game, the L.H.S. does not contain the weakly positive payments at time $t + 1$. In technical terms this means that the stochastic process ${}^N U_{h_t}(\sigma, \tau)$ is a supermartingale. See [11], [15], [21] for references in increasing level of sophistication. The optional stopping time theorem see [11], says that if θ is a stopping time then

$$ {}^N U_{h_t}(\sigma, \tau) \geq E({}^N U_{k_\theta}(\sigma, \tau) | h_t) \tag{A.21} $$

Readers that are not familiar with supermartingales can use the following fact that is actually a slight strengthening of (A.21).

Exercise A.7. *Let* S *be a set of histories such that, for all* $k_s \in S$*,* $h_t \triangleright k_s$ [21] *and, if* $k_s, j_{s'} \in S$*, then* $k_s \triangleright j_{s'} \to k_s = j_{s'}$*. (In other words no history*

[21]this of course implies $s \geq t$

in S is a descendant of another one in S: they are all "sisters", "aunts" or "cousins" with disjoint descendants.) The following inequality holds:

$$ {}^{N}U_{h_t}(\sigma,\tau) \geq \sum_{k_s \in S} {}^{N}U_{k_s}(\sigma,\tau) \tag{A.22}$$

Apart from disregarding the periods from t to s, the weak inequality here is originated by disregarding also payoffs coming from some of descendants of h_t not in S.

B Proofs

Proof of proposition 5.3: We break the proof into two propositions, one shows that are at least one action in $\sigma(h_t)$ and one in $\sigma(\bar{h}_t)$ such that we can increase or decrease their weights without leaving X, the second one uses this fact repeatedly to reduce support and payoffs of σ.

Proposition B.1. *Let $\sigma \in X$, X ESSp, let h_t be an asymmetric history and let supp$(\sigma(h_t)) = \mathbf{B}$, supp$(\sigma(\bar{h}_t)) = \mathbf{C}$.*

Case 1: *If $\mathbf{B}$ has at least two elements and $\mathbf{C}$ is the singleton c there is a pure actions $b \in \mathbf{B}$ such that if $v_- = -\frac{\sigma(h_t)(b)}{1-\sigma(h_t)(b)} < 0$ and if the elementary mutation σ_v is defined for all $v_- \leq v \leq 1$ as:*

$$\sigma_v(k_s) = \begin{cases} vb + (1-v)\sigma(h_t) & \text{if } k_s = h_t \\ \sigma(k_s) & \text{if } k_s \neq h_t, \bar{h}_t \end{cases}$$

then all the σ_v are in X. A similar statement holds if $\mathbf{B}$ is a singleton and $\mathbf{C}$ is not.

Case 2: *Let both $\mathbf{B}$ and $\mathbf{C}$ contain at least two elements. There are two pure actions $b \in \mathbf{B}$ and $c \in \mathbf{C}$ such that if $v_- = -\frac{\sigma(h_t)(b)}{1-\sigma(h_t)(b)}$ and $w_- = -\frac{\sigma(\bar{h}_t)(c)}{1-\sigma(\bar{h}_t)(c)}$ and if the elementary mutation $\sigma_{v,w}$ is defined for all $v_- \leq v \leq 1$ and $w_- \leq w \leq 1$ as:*

$$\sigma_{v,w}(k_s) = \begin{cases} vb + (1-v)\sigma(h_t) & \text{if } k_s = h_t \\ wc + (1-w)\sigma(\bar{h}_t) & \text{if } k_s = \bar{h}_t \\ \sigma(k_s) & \text{if } k_s \neq h_t, \bar{h}_t \end{cases}$$

then all the $\sigma_{v,w}$ are in X.

Note that in either case, the weights of b and c can decrease or increase; the limits on v, w are taken so that σ_v and $\sigma_{v,w}$ assign nonnegative weights on all actions and so are always strategies. At the limits the support becomes strictly smaller.

Proof: First of all, note that, if $p^\sigma(h_t) = 0$ all the σ_v and $\sigma_{v,w}$ are silent mutations and so are in X by proposition 3.2. So we can assume that $p^\sigma(h_t) \neq 0$.

Proof of case 1.

This is easy. Since σ is a Nash equilibrium of ${}^N G$, $\big(\sigma(h_t), \sigma(\bar{h_t})\big)$ is a Nash equilibrium in the forward game ${}^f G^\sigma_{h_t}$ by exercise A.4. So any $b \in \mathbf{B} = supp(\sigma(h_t))$ is a best reply to $\sigma(\bar{h_t})$ and so:

$$
{}^N U(\sigma_v, \sigma) - {}^N U(\sigma, \sigma) = p^\sigma(h_t)[{}^f V^\sigma_{1,h_t}(vb + (1-v)\sigma(h_t), \sigma(\bar{h_t})) - {}^f V^\sigma_{1,h_t}(\sigma(h_t), \sigma(\bar{h_t}))] \equiv 0 \tag{B.1}
$$

by (A.13).

By formula (A.14) the same expression gives ${}^N U(\sigma_v, \sigma_v) - {}^N U(\sigma, \sigma_v)$ and so all σ_v are in X.

Proof of case 2.

Let us choose

$$
(b, c) \in \underset{(x,y), x \in \mathbf{B}, y \in \mathbf{C}}{\text{Argmax}} \; {}^N W_{h_t}(\sigma)(x, y) \subseteq \underset{(\beta,\gamma), \beta \in \Delta(B), \gamma \in \Delta(C)}{\text{Argmax}} \; {}^N W_{h_t}(\beta, \gamma) \tag{B.2}
$$

Since ${}^N W$ is a bilinear function and we are allowed to move β and γ independently of each other, the inclusion of Argmax holds, namely the maximum of ${}^N W_{h_t}(\beta, \gamma)$ is achieved on pure actions. This is the point where we need asymmetry: if h_t where symmetric, we would need to set $\beta = \gamma$ in order get a well defined behavior strategy and the function ${}^N W$, now being quadratic, could have interior, totally mixed, maxima: it is the problem with the Hawk-Dove type of forward game that we saw in the example.

We use the pure actions b, c to construct the $\sigma_{v,w}$ in the statement of the proposition.

It is obvious that σ is an elementary mutation.

As in case 1, since σ is a Nash equilibrium of ${}^N G$, $\big(\sigma(h_t), \sigma(\bar{h_t})\big)$ is a Nash equilibrium in the forward game ${}^f G^\sigma_{h_t}$ by exercise A.4.

We have $b \in \mathbf{B} = supp(\sigma(h_t))$ and $c \in \mathbf{C} = supp(\sigma(\bar{h_t}))$, so b is a best reply to $\sigma(\bar{h_t})$ and c is a best reply to $(\sigma(h_t))$. So ${}^f V^\sigma_{1,h_t}(b, \sigma(\bar{h_t})) = {}^f V^\sigma_{1,h_t}(\sigma, \sigma(\bar{h_t}))$ and ${}^f V^\sigma_{2,h_t}(\sigma(h_t), c) = {}^f V^\sigma_{2,h_t}(\sigma, \sigma(\bar{h_t}))$ which implies that ${}^N U(\sigma_{v,w}, \sigma) =$

${}^NU(\sigma,\sigma)$ by formula (A.16), and so $\sigma_{v,w}$ is a best reply for all allowed values of v and w.

We now check what happens upon meeting the mutants. We have: ${}^NU(\sigma_{v,w},\sigma_{v,w}) - {}^NU(\sigma,\sigma_{v,w}) = p^\sigma(h_t){}^NW_{h_t}^{(\sigma)}(vb+(1-v)\sigma(h_t), wc+(1-w)\sigma(\bar{h}_t))$, by formula (A.17). We call this function $f(v,w)$ and we shall prove that it is identically zero.

We prove first that $f(0,w) = f(v,0) \equiv 0$. If $w=0$ our mutant coincide with σ at $\bar{h}_t$ and so by the remark after Exercise A.2 the expression for $f(v,0)$ coincides with (B.1) and is zero for the same reason.The same holds for $v=0$.

This implies that $\left(\frac{\partial f(v,w)}{\partial v}\right)_{|v=w=0} = \left(\frac{\partial f(v,w)}{\partial u}\right)_{|v=w=0} = \left(\frac{\partial^2 f(v,w)}{\partial^2 v}\right)_{|v=w=0} = \left(\frac{\partial^2 f(v,w)}{\partial^2 u}\right)_{|v=w=0} = 0$.

Now we prove that the mixed derivative $\left(\frac{\partial^2 f(v,w)}{\partial u \partial v}\right)_{|v=w=0}$ is equal to 0. We observe that on one side we have $f(1,1) = p^\sigma(h_t){}^NW_{h_t}^{(\sigma)}(b,c) \geq 0$, because (b,c) is maximizing in equation (B.2). On the other side $f(1,1) = {}^NU(\sigma_{1,1},\sigma_{1,1}) - {}^NU(\sigma,\sigma_{1,1}) \leq 0$ because σ is in an ESSp set and $\sigma_{1,1}$ is a best reply.It follows that $f(1,1)=0$. Together with the vanishing of the other derivatives this suffices to show that f is identically zero, if you are not convinced read the next paragraph.

The function $f(z,z)$ for $0 \leq z \leq 1$ is a quadratic function of one variable that vanish with its first derivative at zero and vanish at one, the only possibility is that it is identically zero. This implies that $\left(\frac{\partial^2 f(v,w)}{\partial u \partial v}\right)_{|v=w=0} = 1/2 \frac{d^2}{dz^2} f(z,z)_{z=0} = 0$. So the function $f(v,w) = {}^NU(\sigma_{v,w},\sigma_{v,w}) - {}^NU(\sigma,\sigma_{v,w})$ is identically zero, because it is at most quadratic and all its first and second derivatives vanish.

And so ${}^NU(\sigma_{v,w},\sigma) = {}^NU(\sigma,\sigma)$ and ${}^NU(\sigma_{v,w},\sigma_{v,w}) = {}^NU(\sigma,\sigma_{v,w})$. But this means that $\sigma_{v,w}$ is in X for all v and w in its range of definition. ■

The next proposition uses the concept of submutation

Proposition B.2. *Let σ be as in the previous proposition, then there is an elementary submutation σ' of σ in X that uses pure actions at h_t and $\bar{h}_t$.*

Proof: Suppose first that both **B** and **C** are larger than a one point set.

We will first show that we can take for σ' one of the $\sigma_{1,1}$, σ_{1,w_-}, $\sigma_{v_-,1}$ and σ_{v_-,w_-} defined in the previous proposition so that it is a submutation.

Let us consider the *linear* function $g(v,w) = {}^N U(\sigma, \sigma_{v,w}) - {}^N U(\sigma,\sigma)$, defined on the rectangle $v_- \leq v \leq 1$ and $w_- \leq w \leq 1$. This function is zero on $(0,0)$, that lies in the interior of the rectangle, so it is either identically zero or, if it is positive on some vertex, it must be negative on some other(s). Choose for σ' a nonpositive one. We now have ${}^N U(\sigma,\sigma') \leq {}^N U(\sigma,\sigma)$, but, by the preceding proposition ${}^N U(\sigma',\sigma') = {}^N U(\sigma,\sigma')$ so ${}^N U(\sigma',\sigma') \leq {}^N U(\sigma,\sigma)$ and σ' is a submutation.

As for the support: when each of these four strategies is applied to h_t and $\bar{h}_t$ it has support strictly smaller than $\mathbf{B}$ and $\mathbf{C}$ respectively:they are either the action b or c or their complements.

If one of the supports is one point you have a segment instead of a rectangle and the proof is the same.

At this point, if σ' does not use already pure actions, we apply repeatedly the previous proposition and the argument above reducing the support by one at each stage until we get the result. $\blacksquare$

This ends the proof of proposition 5.3. $\blacksquare$

Proof of lemma 5.2: We want to estimate

$$\prod_{t=0}^{t=k-1} (1 - \frac{N^\beta - t - 2}{N - t - 2})$$

with $k = N^{1/2} - 1$ and $\beta = 1/2 + \varepsilon$. Some easy algebra shows that $\frac{N^\beta - t - 2}{N-t-2} \geq \frac{N^\beta - N^{1/2}}{N - N^{1/2}}$. So, at each of the k steps up to k, the measure of the set of consistently bad histories will decay at least by the factor $(1 - \frac{N^\beta - N^{1/2}}{N-N^{1/2}}) = (1 - \frac{1-N^{-\varepsilon}}{N^{1/2-\varepsilon} - N^{-\varepsilon}}) \leq (1 - \frac{1}{2N^{1/2-\varepsilon}})$. Where the last inequality holds for $N \geq 2^{1/\varepsilon}$.

So the probability of C_k will be at most

$$(1 - \frac{1}{2N^{1/2-\varepsilon}})^k = [(1 - \frac{1}{2N^{1/2-\varepsilon}})^{N^{1/2-\varepsilon}}]^{\frac{N^{1/2}-1}{N^{1/2-\varepsilon}}}$$

The expression in square brackets converges to $e^{-1/2}$ uniformly in ε and so is smaller than $e^{-1/3}$ for N larger than some N_1, independent of ε. For $N \geq 4$ the exponent $\frac{N^{1/2}-1}{N^{1/2-\varepsilon}}$ is larger than $N^\varepsilon/2$, so if $\bar{N} = max(N_1, 4, 2^{1/\varepsilon})$, the whole expression is less than $exp(-N^\varepsilon/6)$. $\blacksquare$

Proof of theorem 2:

First a trivial remark: if the stage game is doubly symmetric every outcome in it gives equal payoffs to both players. So players have equal payoffs in the repeated game, the forward games and the continuation games. (See A.1 and A.2 for the definitions)

Moreover, if the history h_t is symmetric, the forward game ${}^fG^{\sigma}_{h_t}$ is symmetric, and so it is doubly symmetric, i.e. for h_t symmetric we have ${}^fV_{1,h_t}(x,y) = {}^fV_{2,h_t}(x,y) = {}^fV_{1,h_t}(y,x) = {}^fV_{2,h_t}(x,y)$.

If h_t is asymmetric, ${}^fG^{\sigma}_{h_t}$ is simply a two player game giving equal payoffs to the two players, i.e ${}^fV_{1,h_t}(x,y) = {}^fV_{2,h_t}(x,y)$.

As before, we assume without loss of generality that $P = 1$, $Q = 0$ and so $D = 1$. We let a be an optimal action, so that $u(a,a) = 1$.

We will proceed by induction, for $N = 1$ there is nothing to prove. Suppose the result is true for games repeated $N-1$ times. Let X be an ESSp for ${}^NG$, let $\sigma \in X$. We look at what happens at the zero history.

If *supp*$(\sigma)(\aleph) \neq \mathbf{A}$ we are done by proposition 5.2 with $t = 0$: ${}^NU_{\aleph}(\sigma) \geq N-1$.

If *supp*$(\sigma)(\aleph) = \mathbf{A}$, we prove first an easy lemma.

Lemma B.1. *Let X is an ESSp set for ${}^NG$, $\sigma \in X$ and $p^{\sigma}(h_t) \neq 0$, then the continuation strategy ${}^c\sigma_{h_t}$ is in an ESSp set, ${}^cX$, for ${}^cG_{h_t}$.*

Let ${}^cX$ be the set of restrictions to ${}^cG_{h_t}$ of strategies that agree with σ except after h_t and that are in X. All elementary mutations of these restrictions in ${}^cG_{h_t}$ are also elementary mutation of the original strategies in X and, when playing with elements of ${}^cX$ or among themselves induce the same changes of payoffs, up to the nonzero factor $p^{\sigma}(h_t) \neq 0$.

Now because of full support, (a,a) is a symmetric history with non zero probability, and so, by the lemma, σ restricts to a strategy in an ESSp set in the $N-1$ stages continuation game. This implies, by the induction hypothesis, that ${}^NU_{(a,a)}(\sigma) \geq N-2$.

Since $u(a,a) = 1$ by hypothesis, we have that, in the forward game at stage zero, ${}^fG_{\aleph}$:

$$ {}^fV^{\sigma}_{1,\aleph}(a,a) = u(a,a) + {}^NU_{(a,a)}(\sigma) \geq N-1 \tag{B.3} $$

Let now τ be the elementary mutation playing a with probability 1 at time zero and coinciding with of σ elsewhere. So $\tau(\aleph) = a$, while $\sigma(\aleph)$ will be a full support mixed action, denoted by ξ.

Our result follows from the following sequence of equations:

$$ U(\sigma,\sigma) = {}^fV^{\sigma}_{1,\aleph}(\xi,\xi) = {}^fV^{\sigma}_{1,\aleph}(a,\xi) \tag{B.4a} $$

$$ {}^fV^{\sigma}_{1,\aleph}(a,\xi) = {}^fV^{\sigma}_{1,\aleph}(\xi,a) \tag{B.4b} $$

$$ {}^fV^{\sigma}_{1,\aleph}(\xi,a) \geq {}^fV^{\sigma}_{1,\aleph}(a,a) \tag{B.4c} $$

$$ {}^fV^{\sigma}_{1,\aleph}(\tau,\tau) \geq N-1 \tag{B.4d} $$

Equation (B.4a) holds because $\sigma(\aleph)$ is a Nash equilibrium with full support and so a is a best reply to it.
Equation (B.4b) because the game is a doubly symmetric one.
Equation (B.4c) because of the definition of ESSp.
Equation (B.4d) has been proved from the induction hypothesis in the paragraph above.

■

Proof for example on page 18: We check definition 3.2. Let τ be a strategy in the neighborhood of σ. If τ is to be a best reply to σ, its support must satisfy

$$supp\,(\tau(h_t)) = \begin{cases} \{a,b,c\} & \text{if } h_t = h_0 = \aleph \\ \{b,c\} & \text{if } h_t = (a,a),(b,b) \text{ or } (c,c) \\ c & \text{if } h_t = (a,b),(a,c) \text{ or } (b,c) \\ b & \text{if } h_t = (b,a),(c,a) \text{ or } (c,b) \end{cases} \tag{B.5}$$

so $\tau(h_1) = \sigma(h_1)$ if h_1 is asymmetric. If h_1 is symmetric it is easily seen that $u\,(\sigma(h_1),\tau(h_1)) = 1.5$ and $u\,(\tau(h_1),\tau(h_1)) = 1.5 - \lambda(\tau)$, where $\lambda(\tau) \geq 0$ and strict inequality holds if $\tau \neq \sigma$. It follows that the payoffs of the forward game at stage one are:

$$V^{\sigma,\tau}_{1,\aleph} = V^{\sigma,\sigma}_{1,\aleph} \tag{B.6a}$$
$$V^{\tau,\tau}_{1,\aleph} = V^{\sigma,\sigma}_{1,\aleph} - \lambda(\tau)I \tag{B.6b}$$

where $V^{\sigma,\sigma}_{1,\aleph}$ is the game (4.7) and I denotes the payoffs of the game

$$\begin{array}{cc} & \begin{array}{ccc} a & b & c \end{array} \\ \begin{array}{c} a \\ b \\ c \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

If we call V the matrix of the $V^{\sigma,\sigma}_{1,\aleph}$, and σ_1 and τ_1 the vectors of mixed actions taken at stage one by σ and τ respectively, we have that

$${}^2U(\tau,\tau) - {}^2U(\sigma,\sigma) = (\tau_1 - \sigma_1, V(\tau_1 - \sigma_1)) - \lambda(\tau)\,(\tau_0,\tau_0))$$

by a standard calculation. But now V is negative definite on the orthogonal complement of $(1,1,1)$ and $\lambda(\tau)$ is as above so we have ${}^2U(\tau,\tau) \leq {}^2U(\sigma,\sigma)$ with equality only if $\tau = \sigma$, proving that σ is an ES strategy and so also an ESSet. ■