A Strong Anti-Folk Theorem*

Guilherme Carmona[†]

February 25, 2006

Abstract

We study the properties of finitely complex, symmetric, globally stable, and semi-perfect equilibria. We show that: (1) If a strategy satisfies these properties then players play a Nash equilibrium of the stage game in every period; (2) The set of finitely complex, symmetric, globally stable, semi-perfect equilibrium payoffs in the repeated game equals the set of Nash equilibria payoffs in the stage game; and (3) A strategy vector satisfies these properties in a Pareto optimal way if and only if players play some Pareto optimal Nash equilibrium of the stage game in every stage.

Our second main result is a strong anti-Folk Theorem, since, in contrast to what is described by the Folk Theorem, the set of equilibrium payoffs does not expand when the game is repeated.

Keywords: Nash equilibrium, complexity, stability, social institutions.

*This paper is a revised version of Chapter 3 of my Ph.D. thesis, which has circulated under the title “An Interpretation of Nash Equilibrium Based on the Notion of Social Institutions”. I wish to thank my advisors, Narayana Kocherlakota, and Marcel K. Richter, for their guidance, encouragement, and very helpful comments and suggestions. I wish also to thank an associate editor, an anonymous referee, Kemal Badur, Larry Blume, Nuno Garoupa, Leonid Hurwicz, Max Jodeit, Andy McLennan and participants at the Game 2002, and Asset 2002 conferences for very helpful comments. While they all improved the material in many ways, any remaining errors are mine. Financial support from the *Subprograma Ciência e Tecnologia do 2º Quadro Comunitário de Apoio* is gratefully acknowledged.

[†]Address: Universidade Nova de Lisboa, Faculdade de Economia, Campus de Campolide, 1099-032 Lisboa, Portugal; Phone: (351) 21 380 1671; Fax: (351) 21 387 0933; email: gcarmona@fe.unl.pt.

1 Introduction

The framework of repeated games allows us to understand many familiar social interactions. In particular, as Aumann (1981) pointed out, it can “account for phenomena such as cooperation, altruism, revenge, [and] threats” that would be difficult to rationalize if individuals were to interact just once.

Partly, the importance of these phenomena is that they make efficient outcomes become self-enforcing. This is clearly illustrated by the prisoner’s dilemma: if this game is played infinitely many times (and players are sufficiently patient), then the cooperative outcome is sustained as a subgame perfect equilibrium outcome by the threat of punishing any player that fails to play cooperatively.

Unfortunately, repeating a game also allows for inefficient outcomes to arise as equilibrium outcomes. In fact, the Folk Theorem (see Aumann (1981), Fudenberg and Maskin (1986), and Rubinstein (1979)) shows, under mild conditions, that each individually rational payoff is a subgame perfect equilibrium payoff, provided that the discount factor is sufficiently close to one. Thus, the set of equilibrium payoffs is typically very large, making it hard to predict an actual equilibrium outcome for a given social interaction being analyzed. This creates a serious problem in applying repeated games to real-world problems. In fact, as Aumann (1981) notes, “[i]n a sense, much of the theory of repeated games is an attempt to cut down, in one way or another, the bewildering wealth of equilibrium payoffs provided by [the Folk Theorem].”

We show that, by combining several notions used to study repeated games, the set of equilibrium payoffs in the repeated game becomes equal to the set of Nash equilibrium payoffs in the stage game. This results in a strong anti-Folk Theorem, since, in contrast to what is described by the Folk Theorem, the set of equilibrium payoffs does not expand when the game is repeated.

The notions that we use to define our equilibrium concept are finite complexity, symmetry, global stability and semi-perfection. They were introduced independently by several authors to reflect the view that repeated games strategies represent social institutions, and that these are properties that social institutions typically have, which, in fact, may be necessary for any social institution to endure.

Aumann (1981, p. 21) proposed the use of finite automata in order to

define a simple class of repeated game strategies.¹ A finite automaton strategy encodes the set of all possible histories into a finite set of states, which completely describes players' behavior. Therefore, finite automata strategies are simpler than a general repeated game strategy because these require, in general, an infinite set of states (equal to the countable infinite set of all possible histories) to encode players' behavior.

When players use strategies that can be represented by finite automata, it is natural to define the complexity of a strategy by the size (i.e., the number of states) of the smallest automaton that implements it (see Kalai and Stanford (1988, Theorem 3.1, p. 401).)² To the extent that more complex strategies are more costly to play, players' preferences should depend both on repeated game payoffs and the complexity of the strategies they use. One equilibrium concept defined in terms of such preferences is *semi-perfection* (see Rubinstein (1986, p. 90)). This concept requires not only that each player's strategy maximizes her preferences given the other players' strategies, but also that it remains optimal at all stages of the game.

Symmetry is a property that is often assumed to capture the notion of a common culture or a common legal system. We will call a strategy symmetric if all players share a common state space, a common initial state and a common transition function, properties which were assumed in the notion of social norms used by Okuno-Fujiwara and Postlewaite (1995).

One stability property that has been advocated is that mistakes that players might make in the beginning of the game should not have a long-term effect on players' payoffs — this property was named *global stability* by Kandori (1992, p. 73).³ This property captures a form of forgiveness: if a player fails to choose as dictated by the strategy, one can expect that he will be punished, resulting in a loss in his payoff. However, if the strategy is globally stable, this loss of payoffs is essentially temporary, since eventually his payoff will virtually return to the initial level.

We study the structure of finitely complex, symmetric, globally stable,

¹Finite automata were introduced by McCulloch and Pitts (1943); see also Hopcroft and Ullman (1979). That social institutions should be simple, and represented by finite automata was defended, or at least assumed, by Kandori (1992, p. 72), Okuno-Fujiwara and Postlewaite (1995, p. 83), Rubinstein (1986, p. 84), and Schotter (1981, p. 57).

²Alternative definitions of the complexity of a strategy are proposed in Banks and Sundaram (1990), and in Lipman and Srivastava (1990).

³Some other stability properties that are close in spirit to global stability were proposed by Okuno-Fujiwara and Postlewaite (1995, p. 100).

semi-perfect equilibria. Our main results are:

1. A strategy vector is a finitely complex, symmetric, globally stable, semi-perfect equilibrium only if players play payoff-equivalent Nash equilibria of the stage game in every period.
2. The set of finitely complex, symmetric, globally stable, semi-perfect equilibrium payoffs in the repeated game equals the set of Nash equilibria payoffs in the stage game, and
3. A strategy vector is a Pareto optimal, finitely complex, symmetric, globally stable, semi-perfect equilibrium if and only if players play some Pareto optimal Nash equilibrium of the stage game in every stage.

The first and third results (partially) characterize equilibrium strategies. In both cases, properties of repeated game strategies are directly related to corresponding properties of stage game strategies. While it is clear that repeating a stage game Nash equilibrium forever is a finitely complex, symmetric, globally stable, semi-perfect equilibrium, our first main result shows that this is essentially the only possible case. The possible exception arises when players alternate between different, but payoff-equivalent Nash equilibria of the stage game. In any case, the payoff of such a strategy in the repeated game is the payoff of a Nash equilibrium of the stage game, establishing our second main result.

Furthermore, a strategy in which players alternate between several payoff-equivalent Nash equilibria of the stage game is more complex than one consisting of repeating one such Nash equilibrium, and, clearly, it yields the same payoff to all players. Thus, all players can be better off by playing only one of those Nash equilibria. In other words, the Pareto optimal way of obtaining the properties of finite complexity, symmetry, global stability and semi-perfection requires that players repeat a Pareto optimal Nash equilibrium in every period.

The paper is organized as follows. In Section 2, we introduce our notation and basic definitions. In Section 3, we present our main results. We then discuss several possible interpretations of our results in Section 4. In particular, we discuss how they can provide a rationale for Markovian behavior, an interpretation of Nash equilibria and a framework to address the optimality of social institutions. In Appendix A.1, we establish a converse for our first main result. In Appendix A.2, we study the robustness of our results by considering similar, but different equilibrium concepts.

2 Notation and Definitions

Our notation follows closely the one used by Kalai and Stanford (1988). A *normal form game* G is defined by

$$G = (N, (A_i)_{i \in N}, (u_i)_{i \in N}), \quad (1)$$

where: (1) N is the set of players, (2) A_i is the set of player i 's actions, and (3) for all $i \in N$, $u_i : A \rightarrow \mathbb{R}$, where $A = \prod_{i \in N} A_i$, is player i 's payoff function. We assume that N is a finite set, and that $N = \{1, \dots, n\}$. The set A_i , $i \in N$, may or may not be finite.

The *supergame* of G consists of an infinite sequence of repetitions of G taking place in periods $t = 1, 2, 3, \dots$. At period t the players make simultaneous moves denoted by $a_i^t \in A_i$ and then each player learns his opponent's move.

Finally we need to specify the strategies that players can use and also a way to evaluate payoff in the supergame of G . For all $k \geq 1$, a k -stage history is a k -length sequence $h_k = (a_1, \dots, a_k)$, where, for all $1 \leq t \leq k$, $a_t \in A$; the space of all k -stage histories is H_k , i.e., $H_k = A^k$ (the k -fold Cartesian product of A .) The notation e stands for the unique 0-stage history — it is a 0-length history that represents the beginning of the supergame. The set of all histories is defined by $H = \bigcup_{n=0}^{\infty} H_n$.

For all $h \in H$, define $h^r \in A$ to be the projection of h onto its r^{th} coordinate. For all $h \in H$, we let $\ell(h)$ denote the *length* of h . For any two positive length histories h and $\bar{h}$ in H we define the *concatenation* of h and $\bar{h}$, in that order, to be the history $(h \cdot \bar{h})$ of length $\ell(h) + \ell(\bar{h})$: $(h \cdot \bar{h}) = (h^1, h^2, \dots, h^{\ell(h)}, \bar{h}^1, \bar{h}^2, \dots, \bar{h}^{\ell(\bar{h})})$. We also make the convention that $e \cdot h = h \cdot e = h$ for all $h \in H$.

For all $i \in N$, a *strategy* for player i is a function $f_i : H \rightarrow A_i$ mapping histories into actions. The set of player i 's strategies is denoted by Σ_i , and $\Sigma = \prod_{i \in N} \Sigma_i$ is the joint strategy space. Finally, a strategy vector is $f = (f_1, \dots, f_n)$.

Any strategy $f \in F$ induces a history $h(f)$ as follows:

$$h^1(f) = f(e), \quad h^k(f) = f(h^1(f), \dots, h^{k-1}(f)), \quad (2)$$

for all $k \in \mathbb{N}$.

The payoff in the supergame of G is, for some $\delta \in (0, 1)$, the discounted

sum of stage game payoffs:

$$U^i(f) = (1 - \delta) \sum_{k=1}^{\infty} \delta^{k-1} u^i(h^k(f)). \quad (3)$$

Given an individual strategy $f_i \in \Sigma_i$ and a history $h \in H$, we denote the *individual strategy induced by f_i at h* by $f_i|_h$. This strategy is defined pointwise on H : $(f_i|_h)(\bar{h}) = f_i(h \cdot \bar{h})$, for all $\bar{h} \in H$. We will use $(f|_h)$ to denote $(f_1|_h, \dots, f_n|_h)$ for all $f \in \Sigma$ and $h \in H$. We let $\Sigma_i(f_i) = \{f_i|_h : h \in H\}$ and $\Sigma(f) = \{f|_h : h \in H\}$.

2.1 Automata and Symmetry

For all $i \in N$, let f_i be a strategy for player i . We define an equivalence relation on H , by declaring the histories h and $\bar{h}$ *equivalent relative to f_i* , $h \sim_{f_i} \bar{h}$, if for all histories $\beta \in H$, $f_i(h \cdot \beta) = f_i(\bar{h} \cdot \beta)$. In words, the behavior prescribed by f_i after h is the same as the behavior prescribed by f_i after $\bar{h}$. Let $H / \sim_{f_i}$ denote the quotient set of H relative to $\sim_{f_i}$. Then the complexity of f_i , denoted by $\text{comp}(f_i)$, is defined to be the cardinality of $H / \sim_{f_i}$ (see Kalai (1990, Theorem 4.2, p. 144).) Let F_i denote the set of player i 's strategies with finite complexity, and let $F = \prod_{i \in N} F_i$. Also, let $[h]_{f_i}$ denote the equivalence class of $h \in H$; when it is clear which equivalence relation we are referring to, we write $[h]$ for $[h]_{f_i}$.

An *automaton* for player i is a triple $I_i = ((S_i, s_i^0), T_i, B_i)$ where: S_i is a set of *states*; s_i^0 is an *initial state*; $T_i : S_i \times A \rightarrow S_i$ is a *transition function*; and $B_i : S_i \rightarrow A_i$ is a *behavior function*. Any automaton I_i for player i induces a strategy f_{I_i} as follows: for all histories $h \in H$ with $\ell(h) = m$ define inductively

$$s_i^0(h) = s_i^0, \quad s_i^r(h) = T_i(s_i^{r-1}(h), h^r) \quad \text{for all } r = 1, 2, \dots, m, \quad (4)$$

and then $f_{I_i}(h) = B_i(s_i^m(h))$.

Given a strategy $f_i \in F_i$ for player i , we say that the automaton I_i *implements f_i* if $f_i = f_{I_i}$. For all $f_i \in F_i$ we see that I_{f_i} defined by $S_{f_i} = H / \sim_{f_i}$, $s_{f_i}^0 = [e]$, $T_{f_i}([h], a) = [h \cdot a]$ and $B_{f_i}([h]) = f_i(h)$, implements f_i . Moreover, I_{f_i} is minimal in the sense that if I_i also implements f_i then the cardinality of S_{f_i} is less than or equal to the cardinality of S_i (see Kalai (1990, Theorem 4.2, p. 144), and also Kalai and Stanford (1988, Theorem 3.1, p. 401).) Given this result, we will regard any strategy as an automaton.

The following notation will be useful: Given a strategy $f \in \Sigma$ and a state $s \in S = \prod_{i \in N} S_i$ we denote the *strategy induced by f at s* by $f|s = (f_i|s_i)_{i \in N}$. The strategy $f_i|s_i$ is defined as follows: for all $i \in N$, if $f_i = ((S_i, s_i^0), T_i, B_i)$, then $f_i|s_i$ is defined as the minimal automaton that implements $((S_i, s_i), T_i, B_i)$. Clearly, $\text{comp}(f_i|s_i) \leq \text{comp}(f_i)$ and it may be that $\text{comp}(f_i|s_i) < \text{comp}(f_i)$, as the following example illustrates. Let $\bar{a}_i, \tilde{a}_i \in A_i$, $\bar{a}_i \neq \tilde{a}_i$ and f_i be defined by $S_i = \{\alpha, \beta\}$, $s_i^0 = \alpha$, $T_i(\alpha, a) = T_i(\beta, a) = \beta$ for all $a \in A$, $B_i(\alpha) = \bar{a}_i$ and $B_i(\beta) = \tilde{a}_i$. Then, $f_i|\beta = ((\{\beta\}, \beta), T_{i|\{\beta\} \times A}, B_{i|\{\beta\}})$ and so $\text{comp}(f_i|\beta) = 1 < 2 = \text{comp}(f_i)$.

We will say that a strategy is symmetric if players condition their play on a common state variable, and this common state variable evolves according to a common transition function.

Definition 1 A strategy vector f is **symmetric** if for all $i, j \in N$, $S_i = S_j$, $s_i^0 = s_j^0$, and $T_i = T_j$.

We interpret symmetry as a common culture, or a common legal system. For an example, consider two players who play (a version of) tit-for-tat in a prisoners' dilemma situation. In this case, they all have the same state space ($\{\text{cooperative, non-cooperative}\}$), the same initial state (cooperative) and the same transition function (cooperative follows cooperative if both players play cooperatively, while non-cooperative follows otherwise; and cooperative follows non-cooperative if at least one player plays cooperatively, while non-cooperative follows otherwise). There are many real-world examples that fit this description, for example, an arms race between two countries.

Note that we do not impose that players have the same behavior function in a symmetric strategy. In the above example, some players may choose diverse levels of cooperative and non-cooperative behavior. In the arms race example, a country can respond to an increase in the opponent country's arsenal by also increasing its own arsenal or, alternatively, by invading the other country. In fact, in our notion of symmetry there is no symmetry requirement on the game – players may have different action sets and payoff functions.

2.2 Stability

In this section, we give the stability definition for repeated game strategies that we will use. It is based on the notion of global stability introduced by

Kandori (1992); in fact, it is simply his notion applied to our setting.⁴

Definition 2 *Let $f \in \Sigma$ and let $(\bar{u}_i)_{i \in N} = (U^i(f))_{i \in N}$. Then f is **globally stable** if for all $i \in N$, and $s \in S$,*

$$\lim_{k \rightarrow \infty} U_k^i(f|s) = \bar{u}_i, \quad (5)$$

where $U_k^i(f|s) = (1 - \delta) \sum_{t=k}^{\infty} \delta^{t-k} u_i((f|s)_t)$ is player i 's continuation payoff at stage k .

Note that in the definition of global stability the state s is any possible state, not necessarily the one implied by the play of the equilibrium strategy. The requirement of global stability is that no matter what the past has been, future play of the equilibrium strategies will asymptotically give players the same payoff as if the equilibrium strategies were always followed. Kandori introduced this concept in the context of his work on social norms to capture the intuition that social norms are generally “robust to the mistakes of players” and that they allow “players to test various actions in order to learn [them]”.

2.3 Complexity

As Rubinstein (1998, p. 137) writes,

[a]t the heart of our discussion in this [section] is the trade-off often facing a decision maker when choosing a strategy. On one hand, he hopes his strategy will serve his goals; on the other hand, he would like it to be as simple as possible. There are many reasons why a player may value simplicity: a more complex plan of action is more likely to break down, is more difficult to learn and may require more time to be implemented. We will not examine these reasons here but simply assume that complexity is costly and under the control of the player.

⁴As an anonymous referee pointed out, it might be more reasonable to also impose a restriction on the long-term complexity the strategy. Such a stronger notion of global stability and its implications are developed in Appendix A.2. However, note that all our main results will also hold for that notion simply because it is a stronger concept (see Theorem 5 in Appendix A.2).

We assume as in Abreu and Rubinstein (1988, p. 1264) that players' preferences depend both on repeated game payoffs and the complexity of the strategies they use. For all players $i \in N$, let $\succeq_i$ be player i 's preference relation on $\mathbb{R} \times \mathbb{N}$. We assume that $\succeq_i$

1. is increasing in the payoff:

$$a > b \text{ and } \alpha = \beta \text{ implies } (a, \alpha) \succ_i (b, \beta). \quad (6)$$

2. is decreasing in complexity:

$$a = b \text{ and } \alpha > \beta \text{ implies } (a, \alpha) \prec_i (b, \beta). \quad (7)$$

An equilibrium concept defined using the above preference relation is:

Definition 3 (Rubinstein (1986)) *A strategy vector $f \in F$ is a **semi-perfect equilibrium** if for all $i \in N$, $s \in S$ and $\bar{f}_i \in F_i$ with initial state $\bar{s}_i^0$,*

$$(U^i(f|s), \text{comp}(f_i)) \succeq_i (U^i((\bar{f}_i, f_{-i})|(\bar{s}_i^0, s_{-i})), \text{comp}(\bar{f}_i)). \quad (8)$$

A particular case, considered in Rubinstein (1986), arises when players have lexicographic preferences over payoffs and complexity. This preference relation, denoted by $\succeq_L$ is defined as follows: for all $(a, \alpha), (b, \beta)$ in $\mathbb{R} \times \mathbb{N}$, $(a, \alpha) \succeq_L (b, \beta)$ if and only if $a > b$ or $a = b$ and $\alpha > \beta$. It is then clear that if players have lexicographic preference, any semi-perfect equilibrium is a subgame perfect equilibrium. This result is not true for general preferences in games with more than two players. However, for two-player games with preferences satisfying properties 1 and 2 above, it is true that any semi-perfect equilibrium is a subgame perfect equilibrium (see Piccione (1992, Theorem 2, p. 187)).

3 Structure of Equilibria

In this section, we state and prove our main results, concerning the structure of equilibria. They provide a characterization of the (finitely complex, symmetric, globally stable, semi-perfect) equilibrium payoffs, and of the Pareto optimal equilibria. In addition, they provide necessary conditions on the form of equilibrium strategies.

For all $f \in F$, let $s(f) = \{s(f)_k\}_{k=0}^\infty$ be the sequence of states induced by the play of f and $S(f) = \{s \in S : \text{there exists } k \in \mathbb{N}_0 \text{ such that } s(f)_k = s\}$ be the set of states used by f . If, in addition, f is symmetric, let $D = \{s \in S : s_i = s_j \text{ for all } i, j \in N\}$ be the set of diagonal elements of S . Clearly, $S(f) \subseteq D$.

Lemma 1 below states that for all finitely complex, symmetric, semi-perfect equilibria, all diagonal states will be used in equilibrium, and any diagonal state can be reached starting from any other diagonal state.

Lemma 1 *Let $f \in F$ be a finitely complex, symmetric, semi-perfect equilibrium. Then, for all $s, \bar{s} \in D$, there exists $\{s_k\}_{k=1}^K \subseteq D$ such that $s_1 = s$, $s_K = \bar{s}$ and $s_k = T(s_{k-1}, B(s_{k-1}))$ for all $k = 2, \dots, K$.*

Proof. Let $s, \bar{s} \in D$ and $i \in N$. We claim that if $\bar{s}_i$ can be reached from s , then $\bar{s}$ can be reached from s . Indeed, if $\bar{s}_i$ can be reached from s , then there exists $\{s_k\}_{k=1}^K$ such that $s^1 = s$, $s_i^K = \bar{s}_i$ and $s_k = T(s_{k-1}, B(s_{k-1}))$ for all $k = 2, \dots, K$. Since f is symmetric, then $s_k \in D$ for all $k = 1, \dots, K$. In particular, $s^K = \bar{s}$.

Thus, if $\bar{s} \in D$ could not be reached from $s \in D$, then there is a player $i \in N$ such that $\bar{s}_i$ cannot be reached from s . Thus, player i would deviate from f at state s : he could remove the state $\bar{s}_i$, therefore reducing the complexity and still obtain the same payoff. ■

Lemma 1 implies that we can order D , and thus write $D = \{s_0, s_1, \dots, s_L\}$, with $L = |D| - 1$, in such a way that $s_k = T(s_{k-1}, B(s_{k-1}))$, for all $k = 1, \dots, L$, and $s_0 = T(s_L, B(s_L))$. In words, the play of the game will induce a cycle on the set of diagonal states starting from the first period. Thus, $S(f) = D$.

Since the states that are used in the play of the game will cycle, so will the continuation payoffs for any given player. However, global stability requires that continuation payoffs converge, which is possible only if they are all equal. This is the content of the following lemma.

Lemma 2 *Let $f \in F$ be a finitely complex, symmetric, globally stable, semi-perfect equilibrium, and let $(\bar{u}_i)_{i \in N} = (U^i(f))_{i \in N}$.*

Then for all $i \in N$ and $s \in S(f)$,

$$U^i(f|s) = \bar{u}_i. \quad (9)$$

Furthermore, $u_i(B(s)) = \bar{u}_i$.

Proof. Let $i \in N$ and $s \in S(f)$. Since $D = \{s_0, s_1, \dots, s_L\}$, with $L = |D| - 1$, then $s = s_n$ for some $0 \leq n \leq L$. We have then, for all $k \in \mathbb{N}$,

$$U_k^i(f|s) = \begin{cases} U^i(f|s_n) & \text{if } k = 1, L+1, 2L+1, \dots \\ U^i(f|s_{n+1 \bmod L}) & \text{if } k = 2, L+2, 2L+2, \dots \\ \vdots & \\ U^i(f|s_{n+L \bmod L}) & \text{if } k = L, 2L, 3L, \dots \end{cases} \quad (10)$$

Since by assumption $\lim_{k \rightarrow \infty} U_k^i(f|s) = \bar{u}_i$, it follows that $U^i(f|s_l) = \bar{u}_i$, for all $0 \leq l \leq L$. In particular, $U^i(f|s) = U^i(f|s_n) = \bar{u}_i$.

Finally, since $U^i(f|s) = (1-\delta)u_i(B(s)) + \delta\bar{u}_i$ it follows that $u_i(B(s)) = \bar{u}_i$.

■

As a consequence, we have Theorem 1, which yields a necessary condition on equilibrium strategies.

Theorem 1 *If a strategy $f = ((S, s^0), T, B)$ is a finitely complex, symmetric, globally stable, semi-perfect equilibrium, then $B(s)$ is a Nash equilibrium of G , for all $s \in S(f)$.*

Proof. Let $s \in S(f)$ and $i \in N$. We will show that $B_i(s)$ is a static best reply to $B_{-i}(s)$, which will establish the theorem.

Assume that $B_i(s)$ is not a best reply to $B_{-i}(s)$. Let a^* be a best reply in G against $B_{-i}(s)$. In particular, $u_i(a^*, B_{-i}(s)) > u_i(B(s)) = \bar{u}_i$. Define $\bar{f}_i$ as follows: $\bar{S}_i = S_i$, $\bar{s}_i^0 = s_i$, $\bar{T}_i = T_i$ and

$$\bar{B}_i(\bar{s}_i) = \begin{cases} a^* & \text{if } \bar{s} = s \\ B_i(\bar{s}_i) & \text{otherwise.} \end{cases} \quad (11)$$

It follows immediately that $\text{comp}(\bar{f}_i) \leq \text{comp}(f_i)$. We claim that

$$U_i((\bar{f}_i, f_{-i})|s) > U_i(f|s).$$

Let $s' = T(s, (a^*, B_{-i}(s)))$. Since $\bar{S}_i = S_i$, $\bar{s}_i^0 = s_i$ and $\bar{T}_i = T_i$, it follows that $S(\bar{f}_i, f_{-i}|s') \subseteq D = S(f)$. Thus, player i will receive a payoff of $u_i(a^*, B_{-i}(s)) > \bar{u}_i$ whenever state s is reached after s' and a payoff of $u_i(B(\bar{s})) = \bar{u}_i$ if state $\bar{s} \neq s$ is reached. Hence, $U_i((\bar{f}_i, f_{-i})|s') \geq \bar{u}_i$ (note that it might happen that state s is not reached starting from state s' under $(\bar{f}_i, f_{-i})$). Therefore, it follows that

$$U_i((\bar{f}_i, f_{-i})|s) = (1-\delta)u_i(a^*, B_{-i}(s)) + \delta U_i((\bar{f}_i, f_{-i})|s') > \bar{u}_i = U_i(f|s). \quad (12)$$

This inequality contradicts the fact that f is a semi-perfect equilibrium. Hence, it must be that $B_i(s)$ is a best reply to $B_{-i}(s)$ for all $i \in N$, and so, $B(s)$ is a Nash equilibrium of G . ■

Theorem 1 and Lemma 2 place a strong restriction on the outcomes and payoffs that can arise in equilibrium. They imply that in every stage of a game in which players play in the way described in Theorem 1 we will observe a Nash equilibrium of the stage game. This result yields the following interpretation of Nash equilibria: In a given society, simple, symmetric, stable and minimally complex behavior is described by Nash equilibria.

Let $N(G)$ be the set of Nash equilibrium payoffs in the stage game G , and let $E(G)$ be the set of finitely complex, symmetric, globally stable, semi-perfect equilibrium payoffs in the repeated game. As an immediate consequence of Theorem 1, we obtain the following result.

Theorem 2 *A payoff vector $u = (u_i)_{i \in N}$ is a finitely complex, symmetric, globally stable, semi-perfect equilibrium payoff in the repeated game if and only if u is a Nash equilibrium payoff in the stage game.*

Proof. In symbols, Theorem 2 states that $E(G) = N(G)$. Note first that the strategy of repeating a Nash equilibrium of G forever is a finitely complex, symmetric, globally stable, semi-perfect equilibrium. Thus, $N(G) \subseteq E(G)$.

Conversely, it follows by Theorem 1 and Lemma 2 that if f is a finitely complex, symmetric, globally stable, semi-perfect equilibrium, then $B(s)$ is a Nash equilibrium of G and $u_i(B(s)) = \bar{u}_i$ for all $s \in S(f)$ and all $i \in N$. Thus, $u = (U_i(f))_{i \in N} = (\bar{u}_i)_{i \in N} \in N(G)$. ■

Theorem 2 reinforces the association expressed in Theorem 1 between the Nash equilibrium of a given stage game, and the simple, symmetric, stable, and minimally complex equilibria of its supergame: This association is exact for the payoff that players receive in equilibrium.

Note that the association between Nash equilibrium outcomes of a given stage game and the simple, symmetric, stable, and minimally complex equilibrium outcomes of its supergame is not exact. In fact, we may have an equilibrium outcome in the supergame consisting of oscillations between several different (but payoff equivalent) Nash equilibria.

Consider the following example of a coordination game. Let $n = 2$ and $A_i = \{\alpha_i, \beta_i\}$, $i = 1, 2$. Preferences are lexicographic for both players. Let payoffs be given by Table 1.

$1 \setminus 2$	α_2	β_2
α_1	1, 1	0, 0
β_1	0, 0	1, 1

Table 1: Payoff Function for the Example

Consider the following strategy: $S = \{s^0, s^1\}$, $T(s^0, a) = s^1$, $T(s^1, a) = s^0$ for all a , and $B(s^0) = (\alpha_1, \alpha_2)$, $B(s^1) = (\beta_1, \beta_2)$. This strategy is clearly finitely complex, symmetric, globally stable, and semi-perfect.

However, one may argue that the above equilibrium outcome is unlikely: players would eventually agree to meet in just one location, because they would reduce the complexity of their behavior, while maintaining their payoff. In other words, that strategy is Pareto dominated by another strategy satisfying all the properties we are interested with. In general, we say a strategy σ in a given game is a *Pareto optimal equilibrium* if it is Pareto optimal within the set of all equilibria (i.e., if it cannot be Pareto dominated by any other equilibrium). This definition applies to both the game G and the supergame of G , and for all equilibrium concepts we may wish to use.

Our next result characterizes the set of Pareto optimal, finitely complex, symmetric, globally stable, semi-perfect equilibria: They consist of repetitions of a Pareto optimal Nash equilibrium of the stage game. In this way, we obtain an exact association between Pareto optimal Nash equilibrium *outcomes* of a given game, and the Pareto optimal finitely complex, symmetric, globally stable, semi-perfect equilibrium *outcomes* of its supergame.

Theorem 3 *A strategy f is a Pareto optimal, finitely complex, symmetric, globally stable, semi-perfect equilibrium if and only if $\text{comp}(f_i) = 1$ for all $i \in N$ and $B(s^0)$ is a Pareto optimal Nash equilibrium of G .*

Proof. We start by establishing the if part. The strategy consisting of repeating a Pareto optimal Nash equilibrium of G forever is a finitely complex, symmetric, globally stable, semi-perfect equilibrium. Since $N(G) = E(G)$, then this strategy is also a Pareto optimal equilibrium in the repeated game.

Finally, we turn to the only if part. Let f be a Pareto optimal, finitely complex, symmetric, globally stable, semi-perfect equilibrium. By Theorem 1 and Lemma 2, $B(s)$ is a Nash equilibrium of G and $u_i(B(s)) = \bar{u}_i$ for all

$s \in S$. Since the vector of payoffs $(\bar{u}_1, \dots, \bar{u}_n)$ can be obtained by playing $B(s^0)$ forever, it follows that $\text{comp}(f_i) = 1$ (i.e., if $\text{comp}(f_i) > 1$, then the strategy consisting of repeating $B_i(s^0)$ forever would Pareto dominate f in the repeated game). This, in turn, implies that $B(s^0)$ is a Pareto optimal equilibrium of G (since, otherwise, repeating a Pareto optimal Nash equilibrium of G would Pareto dominate f in the repeated game). ■

4 Discussion

In this section, we discuss our main results by providing alternative interpretations for them.

4.1 Rationalizing Markovian Behavior

We can interpret our main results as stating that static Nash equilibrium is a reasonable concept to describe repeated strategic interactions.

The idea that static Nash equilibria are reasonable descriptions of repeated strategic interactions is also present in the work of Green (1980), Sabourian (1990), and Al-Najjar and Smorodinsky (2001). These authors study repeated interactions in large societies, in which any player's payoffs depend on his choice, and an aggregate outcome. Under different assumptions, they show that in any subgame perfect equilibrium most players play static ε -best replies in every period, i.e., an approximate version of our Theorem 1 holds in their framework. Our results differ from these in two ways: first, although expressing similar ideas, the framework they consider differs considerably from ours; and second, we obtain that, under the assumptions of Theorem 1, exact Nash equilibria will be played.⁵

As pointed out by Al-Najjar and Smorodinsky (2001), we may interpret our results as providing a justification of Markovian behavior. In our framework there is no exogenous state variable, and so in any (time-independent) Markovian strategy players play the same action profile in every period. Thus, Theorem 3 justifies Markovian strategies as a Pareto optimal way of obtaining finite complexity, symmetry, global stability, and semi-perfection.⁶

⁵For an alternative interpretation of Nash equilibrium using large (i.e., non-atomic), but static, games see Barlo and Carmona (2002).

⁶A similar approach is taken in Carmona (2002a, Chapter 4), where monetary trading is rationalized as a Pareto efficient way to obtain the same properties. See also Carmona

Markovian strategies are more appealing when there is an exogenous state variable, i.e., in the stochastic case. Hence, it would be interesting to have an analogue of Theorem 3 for stochastic games. If we allow the initial state of any automata to be randomly determined, then the main result of Carmona (2002a, chapter 4) shows that an analogue of Theorem 3 does not hold — in general, non-Markovian strategies may be needed to obtain the above properties in a Pareto optimal way. However, if the initial state of each automaton is set deterministically (as is the case in the present paper), we conjecture that an analogue of Theorem 3 will hold, at least for a large class of games.

4.2 Interpretations of Nash Equilibrium

In his Ph.D. dissertation, John Nash proposed two interpretations of his equilibrium concept, with the objective of showing how equilibrium points “(...) can be connected with observable phenomenon.” (Nash (1950, p. 21)) One interpretation is rationalistic: if we assume that players are rational, they know the full structure of the game, the game is played just once, and there is just one Nash equilibrium, then players will play according to that equilibrium.

This interpretation was discussed formally by Aumann and Brandenburger (1995), where they identify sufficient epistemic conditions for a Nash equilibrium to be played. One of their conditions is that the conjecture players make about the strategy of the others is commonly known.⁷ This assumption has been criticized by Jacobsen (1996) on the grounds that it is not plausible to assume that one player knows what another player thinks. If we interpret repeated game strategies as social institutions, then any conjecture a player might form about other players can be interpreted as being part of the particular social institution shaping players’ interaction. This interpretation renders Aumann and Brandenburger’s epistemic condition quite reasonable.

A second interpretation of Nash equilibrium, which Nash names mass-action interpretation, is less demanding on the players. In this interpretation, “[i]t is unnecessary to assume that the participants have full knowledge of the

(2002b), in which the same result is established under weaker restrictions in a similar model.

⁷See Aumann and Brandenburger (1995, Theorem B). If there are only two players, then players’ conjectures need only to be mutually known, as their Theorem A shows.

total structure of the game, or the ability and inclination to go through any complex reasoning processes.” (Nash (1950, p. 21)) What is assumed is that there is a population of participants for each position in the game, which will be played throughout time by participants drawn at random from the different populations. If there is a stable average frequency with which each pure strategy is employed by the “average member” of the appropriate population, then this stable average frequency constitutes a Nash equilibrium.

The framework developed by Nash (1950) for the mass-action interpretation is also appealing as a description of individuals’ interaction in large societies. In particular, Okuno-Fujiwara and Postlewaite (1995), and Kandori (1992) have used a similar framework to stress the importance of social institutions in individual decision-making: in environments in which individuals interact throughout time, and have limited information about the others, social institutions can be helpful in summarizing the past, and forming expectations about the behavior of other players.

If we interpret finitely complex, symmetric, globally stable, semi-perfect strategies as enduring social institutions, then we can argue that in those social institutions that are likely to endure, players play a Nash equilibrium in every period. This provides a social institution interpretation of Nash equilibrium.

4.3 On Social Institutions

In this paper, we define social institutions as strategies in some repeated game. Although there may not be complete agreement on the notion of social institutions, several authors have described them in this fashion. For example, Schotter (1981, p. 24) argues that “(...) social institutions can be best described as noncooperative equilibria of supergames that involve repeated play of some particular constituent game (...).”

Viewing social institutions as supergame strategies, several authors have argued that equilibria in such games should satisfy certain additional properties. These additional properties should be imposed not only because they are regarded as desirable from a normative point of view, but also, because they are viewed as properties that social institutions typically have, and which may be necessary for any social institution to endure. In fact, all the properties that we have used have been motivated, at least partly, with such concerns.

If one accepts the view that an equilibrium social institution is, by defin-

ition, a finite complex, symmetric, globally stable, semi-perfect equilibrium, then Theorem 1 describes its structure. Furthermore, with this interpretation in mind, we can also use Theorem 3 to describe the Pareto optimal social institutions in societies that can be appropriately described by a discounted repeated game.

Unfortunately, when applied to the repeated prisoners' dilemma, the above results imply that there is a unique finitely complex, symmetric, globally stable, semi-perfect equilibrium. In this equilibrium players play 'defect' in every period, and so any form of 'cooperative behavior' is impossible to obtain in a finitely complex, symmetric, globally stable and semi-perfect way.

Thus, in social interactions that have a prisoners'-dilemma-like structure, any social institution will produce an inefficient outcome. However, this is in contrast with the view that social institutions (or, at least, those that are optimal) are essentially a way to allow players to obtain a better outcome.

There are at least two ways to address this issue. One way is to add that the properties that we have highlighted only describe decentralized social institutions. With this view, our results would serve as a rationale for government intervention in social interactions with inefficient Nash equilibria, and also to point out that a key issue in such intervention is to make credible the threat of punishments that will not be used in equilibrium.

Alternatively, we can simply take from our results that although each condition that we use seems reasonable by itself, they lead to an implausible conclusion when we combine them.⁸ In this view, we would no longer associate social institutions to finitely complex, symmetric, globally stable, semi-perfect equilibria. Both cases, but especially this one, show that, although repeated games seems to be the natural framework to study social institutions, we still need a good notion of an equilibrium social institution and of an optimal social institution.

A Appendix

A.1 On the Converse of Theorem 1

The following example shows that the condition in Theorem 1 is not sufficient for a strategy to be a finitely complex, symmetric, globally stable, semi-perfect equilibrium. Let $n = 2$ and $A_i = \{\alpha_i, \beta_i\}$, $i = 1, 2$. Preferences are

⁸I thank an associate editor for making this point.

lexicographic for both players. Let payoffs be given by Table 2.

$1 \setminus 2$	α_2	β_2
α_1	$0, 1$	$-1, -1$
β_1	$0, -1$	$0, 1$

Table 2: Payoff Function for the Example

The following strategy f is such that in every stage players play a Nash equilibrium of the stage game, but which is not a semi-perfect equilibrium of the supergame: for all $i = 1, 2$, let $S_i = \{s^0, s^1\}$, $T_i(s^0, a) = s^1$, $T_i(s^1, a) = s^0$ for all $a \in A$, $B_i(s^0) = \alpha_i$ and $B_i(s^1) = \beta_i$. This strategy is clearly finitely complex, symmetric and globally stable. However, f is not semi-perfect: player 1 has an incentive to deviate to $((\tilde{S}_1, \tilde{s}_1^0), \tilde{T}_1, \tilde{B}_1)$ with $\tilde{S}_1 = \{\tilde{s}\}$ and $\tilde{B}_1(\tilde{s}) = \beta_1$.

This example depends crucially on the fact that $u_1(\beta_1, \alpha_2) = u_1(\beta_1, \beta_2)$. However, if we rule out such cases, we can completely characterize the set of finitely complex, symmetric, globally stable, semi-perfect equilibria. Formally, let, for all sets of players N and actions A , $\mathcal{G}(N, A) = \{u = (u_i)_{i \in N} : u_i(a) \neq u_i(\bar{a}) \text{ for all } a, \bar{a} \in A \text{ such that } a \neq \bar{a}\}$ be the set of payoff functions that assign different payoffs to different actions. Clearly, if A is finite, then payoffs in $\mathcal{G}(N, A) \subseteq \mathbb{R}^{n|A|}$ are generic.

The interest in payoff functions belonging to $\mathcal{G}(N, A)$ depends on the following characterization.

Theorem 4 *Let $u \in \mathcal{G}(N, A)$. Then, f is a finitely complex, symmetric, globally stable, semi-perfect equilibrium if and only if, $\text{comp}(f_i) = 1$ for all $i \in N$ and $B(s^0)$ is a Nash equilibrium of G .*

Proof. The if part is obvious, while the only if part follows from Lemma 2 and with Theorem 1. These two results imply that if f is a finitely complex, symmetric, globally stable, semi-perfect equilibrium, then $u_i(B(s)) = u_i(B(s^0))$ for all $s \in S(f)$ and all $i \in N$ and that $B(s^0)$ is a Nash equilibrium of G . Since, $u \in \mathcal{G}(N, A)$, then S_i can have only one element. Thus, $\text{comp}(f_i) = 1$ for all $i \in N$. ■

A.2 Variations

In this appendix, we study whether our results are robust to alternative definitions of stability and minimal complexity.

We can weaken the notion of global stability by requiring that the initial state should not have an effect on the long-run payoff level, without requiring that this payoff level be equal to the equilibrium level. The following definition formalizes this idea.⁹

Definition 4 *Let $f \in \Sigma$. Then f is **weakly globally stable** if for all $i \in N$, and $s, \bar{s} \in S$,*

$$\lim_{k \rightarrow \infty} U_k^i(f|s) = \lim_{k \rightarrow \infty} U_k^i(f|\bar{s}). \quad (13)$$

We can also change the notion of stability by explicitly considering the complexity of the strategy. This is reasonable, since the idea of global stability is that players will eventually be virtually no worse off if some player deviates and players' preferences depend both on the payoff they receive and the complexity of their strategies.¹⁰

Definition 5 *Let $f \in \Sigma$ and let $(\bar{u}_i)_{i \in N} = (U^i(f))_{i \in N}$. Then f is $\succeq$ **globally stable** if for all $i \in N$ and all $s = (s_i)_{i \in N} \in S$,*

$$\lim_{k \rightarrow \infty} (U_k^i(f|s), \text{comp}_k(f_i|s_i)) = (\bar{u}_i, \text{comp}(f_i)), \quad (14)$$

where $\text{comp}_k(f_i|s_i) = \text{comp}(f_i|s_k)$ for all $k \in \mathbb{N}$ and $\{s_k\}_k = s(f_i|s_i)$.

Clearly, we can define the notion of $\succeq$ -weak global stability by appropriately combining the above definitions.

The notion of semi-perfection can also be weakened in a way suggested by Abreu and Rubinstein (1988).

Definition 6 (Abreu and Rubinstein (1988)) *A strategy vector $f \in F$ is a $\succeq$ **-Nash equilibrium** if for all $i \in N$, and $\bar{f}_i \in F_i$,*

$$(U^i(f), \text{comp}(f_i)) \succeq_i (U^i(\bar{f}_i, f_{-i}), \text{comp}(\bar{f}_i)). \quad (15)$$

⁹The definition of weak global stability was suggested to me by Narayana Kocherlakota.

¹⁰This argument and the following definition was suggested to me by an anonymous referee.

One can also define a subgame perfect equilibrium for all supergames in which preferences depend on both payoffs and complexity.

Definition 7 *A strategy vector $f \in F$ is a $\succeq$ -subgame perfect equilibrium if for all $i \in N$, $s \in S$ and $\bar{f}_i \in F_i$ with initial state $\bar{s}_i^0$,*

$$(U^i(f|s), \text{comp}(f_i|s)) \succeq_i (U^i((\bar{f}_i, f_{-i})|(\bar{s}_i^0, s_{-i})), \text{comp}(\bar{f}_i)). \quad (16)$$

It is clear that if f is a $\succeq$ -subgame perfect equilibrium, then f is a $\succeq$ -Nash equilibrium. Furthermore, if f is a finitely complex, symmetric and semi-perfect equilibrium, then f is also a $\succeq$ -subgame perfect equilibrium. This follows from (the proof of) Lemma 1, which implies that $\text{comp}(f_i|s_i) = \text{comp}(f_i)$ for all $s_i \in S_i$ and $i \in N$.

The next result says that weak global stability is all we need for our results.

Theorem 5 *1. If a strategy $f = ((S, s^0), T, B)$ is a finitely complex, symmetric, weakly globally stable, semi-perfect equilibrium, then $B(s)$ is a Nash equilibrium of G , for all $s \in S(f)$.*

2. A payoff vector $u = (u_i)_{i \in N}$ is a finitely complex, symmetric, weakly globally stable, semi-perfect equilibrium payoff in the repeated game if and only if u is a Nash equilibrium payoff in the stage game.

3. A strategy f is a Pareto optimal, finitely complex, symmetric, weakly globally stable, semi-perfect equilibrium if and only if $\text{comp}(f_i) = 1$ for all $i \in N$ and $B(s^0)$ is a Pareto optimal Nash equilibrium of G .

Proof. In order to prove Theorem 5, it is enough to show that Lemma 2 still holds. This is clear: if we define $\bar{u}_i = \lim_{k \rightarrow \infty} U_k^i(f|s)$, $s \in S$, then the same proof can be used. ■

Note that $\succeq$ -global stability is stronger than global stability and that $\succeq$ -weak global stability is stronger than weak global stability. Hence, it follows from Theorem 5 that the conclusion of Theorems 1, 2 and 3 still hold if we replace global stability with either weak global stability, $\succeq$ -global stability or $\succeq$ -weak global stability.

However, our results are sensitive to the notion of equilibrium used. If we use $\succeq$ -Nash equilibrium, then all our results may fail, as the following example below shows.

1\2	C	D	M
C	6, 6	0, 7	0, 0
D	7, 0	1, 53/3	0, 0
M	0, 0	0, 0	9, -1

Table 3: Payoff Function for the Example

Let $n = 2$ and $A_i = \{C, D, M\}$, $i = 1, 2$. Preferences are lexicographic for both players, i.e., $\succeq = \succeq_L$. Let $\delta = 6/10$, and stage game payoffs be given by Table 3.

Consider the following strategy f : for all players $i = 1, 2$, the state space is $S_i = \{M, D, C\}$, and $s_i^0 = M$; the transition function is defined by

$$T_i(M, a) = \begin{cases} D & \text{if } a = (M, M), \\ M & \text{otherwise,} \end{cases} \quad (17)$$

$$T_i(D, a) = \begin{cases} C & \text{if } a = (D, D), \\ D & \text{otherwise,} \end{cases} \quad (18)$$

and

$$T_i(C, a) = \begin{cases} C & \text{if } a = (C, C), \\ M & \text{if } a = (C, D), \\ D & \text{otherwise;} \end{cases} \quad (19)$$

the behavior function is $B_i(M) = M$, $B_i(D) = D$ and $B_i(C) = C$. This strategy is obviously symmetric and has finite complexity. It is also globally stable, since $U^i(f) = 9(1 - \delta) + (1 - \delta)\delta + 6\delta^2 = 6$ and $\lim_{k \rightarrow \infty} U_k^i(f|s) = 6$ for all $s \in S$. It is easy to show that $U_i(f) > U_i(\bar{f}_i, f_{-i})$ for all $i \in N$, and $\bar{f}_i \in \Sigma_i$. Hence, it follows that f is a $\succeq$ -Nash equilibrium of the repeated game.

However, we have that $B(C) = (C, C)$ is not a Nash equilibrium of the stage game; hence the conclusion of Theorem 1 fails. The set of Nash equilibria payoffs in the stage game is $N(G) = \{(1, 1)\}$, but one sees that $U_1(f) = U_2(f) = 6$. Thus, the conclusion of Theorem 2 fails. Finally, the last equality also shows that the strategy that consists of playing the Pareto optimal Nash equilibrium (D, D) of the stage game forever is not Pareto optimal in the repeated game.

Note also that f is $\succeq$ -globally stable since $f_i|M = f_i|C = f_i|D = f_i$ and so $\text{comp}_k(f_i|s_i) = \text{comp}(f_i)$ for all $k \in \mathbb{N}$. Thus, the conclusions of Theorems

1, 2, and 3 may fail when semi-perfection is weakened to $\succeq$ -Nash, even if we strengthen global stability to $\succeq$ -global stability.

Finally, we consider the case where the equilibrium concept is $\succeq$ -subgame perfect equilibrium. For this equilibrium concept, all the results hold if we restrict attention to cycle states. Furthermore, Theorems 2 and 3 still hold with global stability.

For a given symmetric, and finitely complex strategy f let $S_c(f)$ be the set of cycle states. This set is defined as follows: a symmetric, finite automaton induces a sequence of states $\{s_k\}_{k=1}^\infty \subseteq D$, where $s_1 = s^0$, and $s_k = T(s_{k-1}, B(s_{k-1}))$ in a way that $t_2 = \min\{m \in \mathbb{N} : s_{m+1} = s_n \text{ for some } n \leq m\}$ exists. Then $S_c(f) = \{s_{t_1}, \dots, s_{t_2}\}$, where $t_1 \leq t_2$ is such that $s_{t_1} = s_{t_2+1}$.

Lemma 3 *If f is a finitely complex, symmetric, and $\succeq$ -subgame perfect, then $T(s, a)$ belongs to $S_c(f)$ for all $s \in S_c(f)$, and $a \in A$.*

Proof. Let $s \in S_c(f)$, and f be a finitely complex, symmetric, and $\succeq$ -subgame perfect equilibrium. Let $\bar{s} \in T(s, A) = \{\tilde{s} \in S : \text{there exists } a \in A \text{ such that } \tilde{s} = T(s, a)\}$.

Arguing as in Lemma 1, we can show that for all $i \in N$, if $\bar{s}_i$ can be reached from s , then $\bar{s}$ can be reached from s .

Thus, if $\bar{s} \notin S_c(f)$, then $\bar{s}$ cannot be reached from s . Therefore, there is a player $i \in N$ such that $\bar{s}_i$ cannot be reached from s . Thus, player i would deviate from $f_i|s_i$: he could remove the state $\bar{s}_i$, thereby reducing the complexity of $f_i|s_i$ and still obtain the same payoff. But this contradicts the fact that f is $\succeq$ -subgame perfect, and so, we conclude that $\bar{s} \in S_c(f)$. ■

Theorem 6 *If a strategy $f = ((S, s^0), T, B)$ is a finitely complex, symmetric, weakly globally stable, $\succeq$ -subgame perfect equilibrium then $B(s)$ is a Nash equilibrium of G , for all $s \in S_c(f)$. Furthermore, $u_i(s) = u_i(\bar{s})$ for all $i \in N$ and $s, \bar{s} \in S_c(f)$.*

Proof. Assume that there exists $s \in S_c(f) = S(f|s)$ that is not a Nash equilibrium of G . Let $i \in N$, and a_i^* be such that $u_i(a_i^*, B_{-i}(s)) > u_i(B(s))$. Since $T(s, (a_i^*, B_{-i}(s))) \in S_c(f)$ by Lemma 3, and since, as one easily sees, the conclusion of Lemma 2 still holds for all $\tilde{s} \in S_c(f) = S(f|s)$, we can use the proof of Theorem 1 to show that f is not $\succeq$ -subgame perfect. ■

As a consequence, we obtain the following result, stating that the conclusions of Theorems 2 and 3 hold if we replace semi-perfection with $\succeq$ -subgame perfection.

Theorem 7 1. A payoff vector $u = (u_i)_{i \in N}$ is a finitely complex, symmetric, globally stable, $\succeq$ -subgame perfect equilibrium payoff in the repeated game if and only if u is a Nash equilibrium payoff in the stage game.

2. A strategy f is a Pareto optimal, finitely complex, symmetric, globally stable, $\succeq$ -subgame perfect equilibrium if and only if $\text{comp}(f_i) = 1$ for all $i \in N$ and $B(s^0)$ is a Pareto optimal Nash equilibrium of G .

Proof. Abusing notation, let $E(G)$ denote the set of finitely complex, symmetric, globally stable, $\succeq$ -subgame perfect equilibrium payoffs in the repeated game.

Let $f \in F$ be a finitely complex, symmetric, globally stable, $\succeq$ -subgame perfect equilibrium and $s \in S_c(f)$. Then, $\lim_{k \rightarrow \infty} U_k^i(f|s^0) = u_i(s) \in N(G)$ by Theorem 6. Since f is globally stable, then $U^i(f) = u_i(s)$ and so $E(G) \subseteq N(G)$.

The remainder of the proof follows the same arguments used in the proofs of Theorems 2 and 3. ■

Furthermore, if we strengthen global stability to $\succeq$ -global stability, then the conclusions of Theorems 1, 2, and 3 still hold when semi-perfection is replaced by $\succeq$ -subgame perfection.

Theorem 8 1. If a strategy $f = ((S, s^0), T, B)$ is a finitely complex, symmetric, $\succeq$ -globally stable, $\succeq$ -subgame perfect equilibrium, then $B(s)$ is a Nash equilibrium of G , for all $s \in S(f)$.

2. A payoff vector $u = (u_i)_{i \in N}$ is a finitely complex, symmetric, $\succeq$ -globally stable, $\succeq$ -subgame perfect equilibrium payoff in the repeated game if and only if u is a Nash equilibrium payoff in the stage game.

3. A strategy f is a Pareto optimal, finitely complex, symmetric, $\succeq$ -globally stable, $\succeq$ -subgame perfect equilibrium if and only if $\text{comp}(f_i) = 1$ for all $i \in N$ and $B(s^0)$ is a Pareto optimal Nash equilibrium of G .

Proof. By Theorems 6, 9, and 10, it is enough to show that $S(f) = S_c(f)$.

Let $s \in S_c(f)$. Then, $\text{comp}(f_i|s_i) = |S_c(f)|$ by Lemma 3. Thus, since $\lim_{k \rightarrow \infty} \text{comp}_k(f_i|s_i^0) = \text{comp}(f_i|s_i)$, and since f is $\succeq$ -globally stable, it follows that $\text{comp}(f_i) = |S_c(f)|$. Since $S_c(f) \subseteq S(f)$, it follows that $S(f) = S_c(f)$. ■

In contrast to Theorems 7 and 8, we find that the conclusions of Theorems 1, 2, and 3 may fail if we weaken the notion of stability from $\succeq$ –global stability to weak global stability. However, they still hold if we restrict ourselves to the cycle state, as Theorems 9 and 10 below show.

Let $E_c(G)$ be the cycle payoff vectors supported by finitely complex, symmetric, weakly globally stable, $\succeq$ –subgame perfect equilibria. These payoff vectors can be thought of as the payoff players will obtain in the long run (note that these payoffs are well defined by an analogous version of Lemma 2.) As a consequence of Theorem 6 we have that $E_c(G) = N(G)$.

Theorem 9 *A payoff vector $u = (u_i)_{i \in N}$ is a cycle payoff vector supported by a finitely complex, symmetric, weakly globally stable, $\succeq$ –subgame perfect equilibrium payoff in the repeated game if and only if u is a Nash equilibrium payoff in the stage game.*

Finally, in order to obtain an analog to Theorem 3, we define the following notion of long-run Pareto optimality: an equilibrium $f \in F$ is *Pareto optimal in the long run* if it is not Pareto dominated in the long run by any other equilibrium. An equilibrium $f \in F$ *Pareto dominates an equilibrium g in the long run* if for all $s \in S_c(f)$,

$$(U_i(f|s), \text{comp}(f_i|s_i)) \succeq (U_i(g|s), \text{comp}(g_i|s_i)), \quad (20)$$

for all $i \in N$, and

$$(U_j(f|s), \text{comp}(f_j|s_j)) \succ (U_j(g|s), \text{comp}(g_j|s_j)), \quad (21)$$

for at least one $j \in N$. We then obtain

Theorem 10 *A strategy $f = ((S, s^0), T, B)$ is a Pareto optimal, finitely complex, symmetric, weakly globally stable, $\succeq$ – subgame perfect equilibrium if and only if $\text{comp}(f_i|s_i) = 1$ for all $i \in N$ and $B(s^0)$ is a Pareto optimal Nash equilibrium of G .*

We can give an example that shows that we cannot extend the conclusions of Theorems 6, 9, and 10 to all states if we use either weak global stability or $\succeq$ –weak global stability. The example is like the previous one, except that $u_1(M, M) = 3 = u_2(M, M)$. Let $n = 2$ and $A_i = \{C, D, M\}$, $i = 1, 2$. Preferences are lexicographic for both players. Let $\delta = 6/10$, and stage game payoffs be given by Table 4.

1\2	C	D	M
C	6, 6	0, 7	0, 0
D	7, 0	1, 1	0, 0
M	0, 0	0, 0	3, 3

Table 4: Payoff Function for the Example

Consider the following strategy f : for all players $i = 1, 2$, the state space is $S_i = \{s^0, s^1, s^2\}$; the transition function is

$$T_i(s^0, a) = \begin{cases} s^1 & \text{if } a = (C, C), \\ s^2 & \text{otherwise,} \end{cases} \quad (22)$$

$T_i(s^1, a) = s^2$, and $T_i(s^2, a) = s^2$ for all $a \in A$; the behavior function is $B_i(s^0) = C$, $B_i(s^1) = M$ and $B_i(s^2) = D$. This strategy is obviously symmetric and has finite complexity. It is weakly globally stable since $\lim_{k \rightarrow \infty} U_k^i(f|s) = 1$ for all $s \in S$. It is easy to show that $U_i(f|s^l) > U_i(\bar{f}_i, f_{-i} | (\bar{s}_i^0, s_{-i}^l))$ for all $i \in N$, $s^l = s^0, s^1, s^2$, and $\bar{f}_i \in \Sigma_i$ with initial state $\bar{s}_i^0$. Hence, it follows that f is a $\succeq$ -subgame perfect equilibrium of the repeated game.

However, we have that $B(s^0) = (C, C)$ is not a Nash equilibrium of the stage game; hence the conclusion of Theorem 1 fails. The set of Nash equilibrium payoffs in the stage game is $N(G) = \{(1, 1), (3, 3)\}$, but one sees that $U_1(f) = U_2(f) = 6(1 - \delta) + 3\delta(1 - \delta) + \delta^2 > 3$. Thus, the conclusion of Theorem 2 fails. Finally, the last inequality also shows that the strategy that consists of playing the Pareto optimal Nash equilibrium (M, M) of the stage game forever is not Pareto optimal in the repeated game.

Note also that f is $\succeq$ -weakly globally stable since $\lim_{k \rightarrow \infty} U_k^i(f|s) = u_i(D, D)$ for all $s \in S$ and $\lim_{k \rightarrow \infty} \text{comp}_k(f_i|s_i) = \text{comp}(f_i|D) = 1$ for all $s_i \in S_i$ (since $f_i|D = ((\{D\}, D), T_{i|\{D\} \times A}, B_{i|\{D\}}))$). Thus, we cannot extend the conclusions of Theorems 6, 9, and 10 to all states even if we strengthen weak global stability to $\succeq$ -weak global stability.

Finally, we show that the conclusion of Theorem 1 does not hold if we use $\succeq$ -subgame perfection together with global stability. Again, let $n = 2$, $\delta = 6/10$ and $A_i = \{C, D, M\}$, $i = 1, 2$. Preferences are again lexicographic for both players. The stage game payoffs are changed and are given by Table 5.

$1 \backslash 2$	C	D	M
C	6, 6	0, 7	-8, -8
D	7, 0	1, 1	-8, -8
M	-8, -8	-8, -8	$-22/3, -22/3$

Table 5: Payoff Function for the Example

Then, f , as defined in the previous example, is globally stable since $U^i(f) = 6(1 - \delta) - \frac{22}{3}(1 - \delta)\delta + (1 - \delta)\delta^2 = 1$. Since $B(C) = (C, C)$ is not a Nash equilibrium of the stage game, the conclusion of Theorem 1 fails.

The following table summarizes the results obtained for $\succeq$ -subgame perfection.

	Theorem 1	Theorem 2	Theorem 3
global stability	C	Y	Y
weak global stability	C	C	C
$\succeq$ -global stability	Y	Y	Y
$\succeq$ -weak global stability	C	C	C

Table 6: Results for $\succeq$ -subgame perfection

The symbol Y means that the conclusion of the corresponding Theorem holds. The symbol C means that the conclusion of the corresponding Theorem holds only for cycle states.

Another variation that can be made is in the definition of complexity of a strategy. Banks and Sundaram (1990), in the same framework as ours but with only two players, consider a measure of complexity that takes into account not only the number of states of the smallest automaton implementing it, but also the number of transitions emanating from each state. Using this complexity measure, they show that assuming finite complexity is enough to obtain the conclusion of Theorem 1. However, analogues of Theorem 2 and 3 are, in general, false.

As an example, consider the battle of the sexes game (see Fudenberg and Tirole (1991, page 19)) in which only pure actions are allowed, and in which players have lexicographic preferences in its supergame. Then, alternating between the two Nash equilibria of the stage game is a Pareto optimal equilibrium in the sense of Banks and Sundaram (1990), and the

resulting supergame payoff does not equal the payoff of any stage game Nash equilibria.

References

- ABREU, D., AND A. RUBINSTEIN (1988): “The Structure of Nash Equilibrium in Repeated Games with Finite Automata,” *Econometrica*, 56, 1259–1281.
- AL-NAJJAR, N., AND R. SMORODINSKY (2001): “Large Nonanonymous Repeated Games,” *Games and Economic Behavior*, 37, 26–39.
- AUMANN, R. (1981): “Survey of Repeated Games,” in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*. Bibliographisches Institut, Mannheim.
- AUMANN, R., AND A. BRANDENBURGER (1995): “Epistemic Conditions for Nash Equilibrium,” *Econometrica*, 63, 1161–1180.
- BANKS, J., AND R. SUNDARAM (1990): “Repeated Games, Finite Automata, and Complexity,” *Games and Economic Behavior*, 2, 97–117.
- BARLO, M., AND G. CARMONA (2002): “Strategic Behavior in Non-Atomic Games: A Refinement of Nash Equilibrium,” mimeo, University of Minnesota, and Universidade Nova de Lisboa.
- CARMONA, G. (2002a): “Essays on the Theory of Social Institutions, with Special Emphasis on Monetary Trading,” Ph.D. thesis, University of Minnesota.
- (2002b): “Monetary Trading: An Optimal Exchange System,” Working Paper 420, Universidade Nova de Lisboa.
- FUDENBERG, D., AND E. MASKIN (1986): “The Folk Theorem in Repeated Games with Discounting and Incomplete Information,” *Econometrica*, 54, 533–554.
- FUDENBERG, D., AND J. TIROLE (1991): *Game Theory*. The MIT Press, Cambridge, MA.

- GREEN, E. (1980): “Non-Cooperative Price Taking in Large Dynamic Markets,” *Journal of Economic Theory*, 22, 155–182.
- HOPCROFT, J., AND J. ULLMAN (1979): *Introduction to Automata Theory, Language, and Computation*. Addison-Wesley, New York.
- JACOBSEN, H. (1996): “On the Foundations of Nash Equilibrium,” *Economics and Philosophy*, 12, 67–88.
- KALAI, E. (1990): “Bounded Rationality and Strategic Complexity in Repeated Games,” in *Game Theory and Applications*, ed. by A. N. Tatsuuro Ichiishi, and Y. Tauman. Academic Press, New York.
- KALAI, E., AND W. STANFORD (1988): “Finite Rationality and Interpersonal Complexity in Repeated Games,” *Econometrica*, 56, 397–410.
- KANDORI, M. (1992): “Social Norms and Community Enforcements,” *Review of Economic Studies*, 59, 63–80.
- LIPMAN, B., AND S. SRIVASTAVA (1990): “Informational Requirements and Strategic Complexity in Repeated Games,” *Games and Economic Behavior*, 2, 273–290.
- MCCULLOCH, W., AND W. PITTS (1943): “A Logical Calculus for the Ideas Immanent in Nervous Activity,” *Bulletin of Mathematical Biophysics*, 5, 115–133.
- NASH, J. (1950): “Non-Cooperative Games,” Ph.D. thesis, Princeton University.
- OKUNO-FUJIWARA, M., AND A. POSTLEWAITE (1995): “Social Norms and Random Matching Games,” *Games and Economic Behavior*, 9, 79–109.
- PICCIONE, M. (1992): “Finite Automata Equilibria with Discounting,” *Journal of Economic Theory*, 56, 180–193.
- RUBINSTEIN, A. (1979): “Equilibrium in Supergames with the Overtaking Criterion,” *Journal of Economic Theory*, 21, 1–9.
- (1986): “Finite Automata Play the Repeated Prisoner’s Dilemma,” *Journal of Economic Theory*, 39, 83–96.

——— (1998): *Modeling Bounded Rationality*. The MIT Press, Cambridge, MA.

SABOURIAN, H. (1990): “Anonymous Repeated Games with a Large Number of Players and Random Outcomes,” *Journal of Economic Theory*, 51, 92–110.

SCHOTTER, A. (1981): *The Economic Theory of Social Institutions*. Cambridge University Press, Cambridge.